

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

## *Collocations in Science Writing*

Christopher Gledhill (2000).

Language in Performance Series No. 22, Tübingen: Gunter Narr  
Verlag

ISBN 3-8233-4945-7.

### **3. Definitions of Collocation**

A collocation is a familiar recurrent expression. For many linguists, collocations are related to a range of commonly recognised multi-word phrases in language, including catchphrases, clichés, fixed expressions, formulae, free and bound collocations, idioms, lexical phrases, turns-of-phrase and so on. Collocation has been defined in various ways, and definitions depend on the specific aims of the observer. Phraseologists and dictionary makers, for example, examine the way lexical words behave in certain combinations. The adjectives *strong* and *powerful* can thus be seen to have a similar meaning but a different range of use with certain nouns: *strong argument*, *powerful argument* versus *strong tea* / *\*powerful tea*, *\*strong car* / *powerful car*. Once such a restriction is identified for a pair of words, we are dealing with some form of collocation. However, as the word ‘familiar’ suggests in my working definition, there is more to collocation than the combination of two or more words. In the following discussion, I attempt to synthesise three different ways of categorising and defining the notion of collocation: Halliday’s *statistical / textual* view, the *semantic / syntactic* tradition in lexicology, and the *discoursal / rhetorical* model from discourse analysis. I then go on to propose an overall model of phraseology which serves as a basis for the analysis carried out in the rest of the book. In the corpus analysis sections of this book, Halliday’s statistical definition is specifically taken as the first and simplest stage of my analysis, but is then supplemented by further stages of interpretation in order to determine the structural and rhetorical significance of the collocations identified in the corpus.

From a **statistical / textual** perspective, it is generally agreed that no one linguistic definition of collocation is entirely reliable when it comes to

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

finding expressions systematically in large numbers of texts. For this practical reason, collocations have often been defined statistically in corpus-based studies, especially if the analyst is attempting to find examples of typical style. The first stage of analysis to be used in this book therefore follows Halliday, who frames collocation in terms of statistical probability and co-occurrence:

Collocation is the syntagmatic association of lexical items, quantifiable, textually, as the probability that there will occur at  $n$  removes (a distance of  $n$  lexical items) from an item  $x$ , the items  $a, b, c \dots$ . Any given item thus enters into a range of collocation, the items with which it is collocated being ranged from more to less probable. (Halliday 1961:276).

Van Roey summarises this view in terms of expression or 'usage':

[collocation is] that linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than its 'synonyms' because of constraints which are not on the level of syntax or conceptual meaning but on that of usage. (van Roey 1990:46).

A collocate can thus simply be seen as any word which co-occurs within an arbitrarily determined distance or *span* of a central word or *node*. Collocation is thus considered to be the frequency with which collocates co-occur with one node relative to their frequency of collocation with other nodes. From the point of view of many corpus linguists, all that separates collocation from mere word co-occurrence is the statistical level at which the researcher is happy to say that the co-occurrence is not accidental. This approach is also 'textual' in that it relies solely on the ability of the computer program to analyse large amounts of computer-readable texts. Sinclair (1991:68) shows this by noting that the independent probability of 'set' collocating with 'off' in the Cobuild corpus is just one in a million (1 855 instances of 'set' multiplied by 556 instances of 'off' from a total of 7.3 million words). Yet the actual frequency of collocation is around 550 instances (that is: 70 in a million). The expression 'set off' can thus be considered a significant collocation without considering other semantic or lexical considerations (1987b:153).

This perspective essentially emphasises collocation as co-occurrence (words which frequently combine) and recurrence (combinations which frequently occur in language). The notion of statistical collocation is integral to Halliday's theory of discourse and the theory is discussed in section III. It is sufficient to note here that a statistical view of language allows the linguist

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

to identify patterns that would not normally be recognised using traditional categories. The textual view of collocation also emphasises the fact that collocations are not disembodied lexical units inserted into the body of a text without modification, but are the result of reformulations and paraphrases which have developed throughout the length of a text. A textual collocation is likely to have a specific textual function or may occur in a rather restricted set of contexts. These expressions can be seen to be couched seamlessly in the surrounding text, and in many of the examples we see below, the collocational patterns of a specific phrase are motivated or triggered by other phrases which appear to be at some distance (a phenomenon observed by Phillips 1985 and Hoey 1991). This is what is meant by ‘long-range collocation’.

In contrast, the **semantic / syntactic** tradition defines collocation as a more abstract relationship between words, without reference to frequency of occurrence or probability, shifting the emphasis therefore from the textual co-occurrence of an expression to its potential for lexical combinability. While Halliday’s approach to collocation is appropriate to a discussion of discourse and register, style is not the main concern in lexicology. Instead the emphasis is on dictionary making and terminology, and collocations are typically seen either as units of meaning (lexical items or idioms) or units of grammar (phrases). It is for this reason that collocation is usually seen as a rather restricted category of expression and is also typically limited to the lexical relation between content words. The standard definition is given by Benson:

Collocations ... are fixed recurrent combinations of words in which each word basically retains its meaning. (Benson 1989:85).

Howarth (1996) has presented a synthesis of the mainstream ideas of lexicology and phraseology studies, taking particular account of the Russian perspective (Dobrovol’skij 1992). He notes that the ‘composite unit’ is traditionally classified according to two measures (1996: 36-46):

‘Commutability’ - The extent to which the elements in the expression can be replaced or moved. As in the free collocation *make a decision* where *make* can be replaced by a series of de-lexical verbs *reach, take* etc., while in the restricted collocation *shrug one’s shoulders* there is no alternative to the verb *shrug*.

‘Motivation’ - The extent to which the semantic origin of the expression is identifiable, as in the figurative idiom *move the goalposts* [to change the

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

required conditions for success], as opposed to the opaque idiom *shoot the breeze* [to chatter].

Fixed expressions are characterised by the relationship between their component words and the overall meaning of the phrase. Cruse (1986) thus distinguishes collocation as ‘syntagmatically simple’ i.e. an expression composed of one word in its normal sense with another restricted word (as in: *table a resolution, tender one’s resignation*) and idiom as ‘semantically simple’ i.e. as a single choice of meaning with an unpredictable or non-compositional sequence of words (*let the cat out of the bag, spill the beans*). In Howarth’s lexical continuum model (1996:32-33), collocations are placed on a sliding scale of meaning and form from relatively unrestricted (collocations) to highly fixed (idioms):

Free collocation	<i>blow a trumpet</i>	‘to play the trumpet’
Restricted collocation	<i>blow a fuse</i>	‘to destroy a fuse’, or (idiomatic) ‘get angry’
Figurative idiom	<i>blow your own trumpet</i>	‘to boast, sell oneself excessively’
Pure idiom	<i>blow the gaff</i>	‘to reveal a concealed truth’

The problem commonly encountered with these classifications (as can be seen in the ambiguous example of *to blow a fuse*) is that is difficult to determine what is meant by ‘syntactically fixed’, ‘unmotivated’ or ‘opaque’.

In addition to the notion of the collocational continuum, one of the most influential ideas to emerge from the field of lexicography involves Mel’čuk’s theory of lexical functions. Mel’čuk defines collocation as an semantic function operating between two or more words in which one of the words keeps its ‘normal’ meaning (Mel’čuk 1995:182). Fontenelle explains this abstract relationship:

[...] the concept of collocation is independent of grammatical categories: the relationship which holds between the verb *argue* and the adverb *strongly* is the same as that holding between the noun *argument* and the adjective *strong*. (Fontenelle 1994:43).

For example, several restricted collocations in English have the abstract function of ‘intensifier’ (coded by Mel’čuk as ‘*magn*’): *stark naked, utter foolishness, piping hot*. The vocabulary as a whole is therefore organised into a grammar of intensity, of quantity (*a speck of dust, a pride of lions*), of

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

operation (*to lend support, to deal a blow*), of function (*war is raging, silence reigns*) and so on (Mel'čuk 1998:36-41). By bringing disparate collocational patterns into a broad theory of meaning, Mel'čuk has argued for a universal typology of lexical functions which are realised by a delimited number of underlying lexical functions in English and other languages.

In lexicology and phraseology studies, idioms are seen as the prime examples of semantic and syntagmatic units, and have a correspondingly privileged status (Howarth 1998:169). On the other hand, collocations emerge as less tidy and easy to categorise, being seen as increasingly less fixed and also more diffuse – largely of course because they are often defined in terms that make idioms generally appear to be ideal units. Collocations also tend to be defined as a subcategory of other items. Mel'čuk, for example, sees them as a very specific category: 'Collocations – no matter how one understands them – are a subclass of what are known as *set phrases*' (Mel'čuk 1998:23). Approaching the issue from a different perspective, van der Wouden (1997) has argued that collocation should be seen as the central term in lexicology. He points out that regardless of the way collocations are defined, analysts find more instances of collocation than of idiom in actual texts, and proposes that the notion of 'collocability' requires better definition than the more peripheral idea of 'idiomaticity'. Like many linguists in the generative field (for example, Abeillé 1995), he sees syntagmatic variability as key to the notion of a fixed expression, and suggests that many features of language are idiomatic in this sense:

I will use the term COLLOCATION as the most general term to refer to all types of fixed combinations of lexical items. In this view idioms are a special subclass of collocations, to wit those collocations with a non-compositional, or opaque semantics. An idiom might even be defined as any grammatical form whose meaning is not deducible from its structure. In this view all morphemes are idioms. (van der Wouden 1997:9).

Makkai (1992) has similarly argued that collocations and idioms can be seen as extended forms of words. Kjellmer makes a similar point:

Highly distinctive collocations behave in important respects like one-word lexemes. They are often semantically identical or almost identical with single words. (Kjellmer 1984)

Van der Wouden further makes the point that idioms and collocations share a number of properties, not least of which the ability to contain analogies which are not carried on into the rest of the language system:

[...] you cannot predict that the meaning of *sleep like a log* will denote an intense form of sleeping, but after you have learned what it means, you see that *like a log* is an intensifier. The essence of collocation is that the assignment of *like a log* to the meaning 'very' does not feed other combinations. So even though we have a meaning for it, that meaning is only valid in a certain collocation [...] (van der Wouden 1997:54-55).

From this discussion, it emerges that the distinction between idiom and collocation is difficult to justify on purely semantic or syntagmatic grounds. Instead, collocation constitutes a general system of abstract relations which underpin much phraseology in the language, and range from relatively free to relatively fixed expression. A different perspective, although still within our 'semantic / syntactic' framework, relates collocational patterns to the wider grammatical system, as in the work of Sinclair (1991). For example, Renouf and Sinclair (1991) have noted that the meaning of a lexical item can be predicted by the presence of grammatical items and the sequence in which they are arranged. Thus in expressions such as *an X of*, X is often a quantity, or in *too Y in the Z*, Y and Z are often time expressions (such sequences are termed collocational frameworks). Louw (1993) has noted that clusters of lexical collocations often share a similar semantic profile or 'semantic prosody'. Thus the NP subjects of the phrasal verb *set in* belong invariably to a semantic field with negative associations (*the bad weather, gangrene, the rot, depression ... sets in*). According to this perspective, the grammatical patterns of co-occurrence are an intrinsic meaning of an expression, and any item which is inserted into the pattern can be re-interpreted in terms of the existing collocational framework (e.g. *a cacophony of musicians* [collective], *the Labour party have set in* [negative connotation]).

In a large-scale study of verb complementation, Hunston and Francis (1998) similarly make a specific link between the grammatical form of an expression (its underlying word class pattern) and its meaning, claiming that the pattern is part of the meaning of the expression. Hunston and Francis identify a number of collocations which share specific grammatical patterns and yet also display a closely related meaning. Here is one example:

...sense and pattern tend to be associated with each other, such that a particular sense of a verb may be identified by its pattern. The verb *recover* has two main senses: 'to get better' following an illness or period of unhappiness, and 'to get back' something that was lost. The first of these senses has the pattern 'V from n' (e.g. *He is recovering from a knee injury*) [...] and 'V' (e.g. *It took her three days to recover*), whilst the second has

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

the pattern 'V n' (e.g. *Police... recovered stolen goods*). (Hunston and Francis 1998:51).

This can be seen to be an extension of the general principle of delexicalisation, in which lexical items merge into grammatical forms, effectively becoming grammatical collocations (grammatical words collocating with lexical words). The expressions created by grammatical collocation and colligation depend in turn on a notion of extended meaning, as argued by Renouf (1998). The extended meaning of a word or expression is built up over time by its collocational tendencies within different texts. Thus while lexicologists conceive of collocation as a lexical unit and examine the behaviour of component words within this larger lexical item, Firthian and Hallidayan linguists see collocation as a specific grammatical pattern, associated with a particular meaning. The work of Louw, Renouf, Hunston, Francis and others has been much influenced by Sinclair's notion of the 'idiom principle'. Sinclair (1991) argued that meaning is organised through language not by filling lexical items into grammatical context-free slots, but in a system where structure maps onto meaning very closely. He emphasises the importance of syntagmatic sequences as single functional choices, and argues that neither individual words nor deep syntactic structures correspond to natural choice in language:

The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. To some extent, this may reflect the recurrence of similar situations in human affairs; it may illustrate a natural tendency to economy of effort or it may be motivated in part by the exigencies of real-time conversation. (Sinclair 1987c: 320)

From the 'semantic / syntactic' perspective, we have seen that the notion of collocation has been extended from traditional restricted collocations and idioms (*curry favour, strike a chord*) to less conventional notions such as grammatical collocation (linking grammatical items with lexical items, as in phrasal verbs *refer to, answer for*) and de-lexical verbs (*have a break, take a decision*). Many of these patterns can be seen to obey underlying lexical relationships. The notion has recently been applied to a much wider category of expression following work in corpus analysis, including semantic prosody (clusters of semantically related words: *push through [a reform, a project, a law...]*), collocational frameworks (lexical and grammatical collocation: *not only... but also, find / make it [easy, difficult, hard, impossible] to + clause*)

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

and colligation (collocation between grammatical categories, e.g. the set of nouns that can introduce NP complement clauses: *the idea, conviction, belief, thought that*). These patterns demonstrate the close correlation between syntax and semantics and are seen as a confirmation of Halliday's (1985) notion of a lexico-grammar: a theory of lexis and grammar as an interrelated continuum rather than as separate levels.

So far we have seen collocations as 'statistical / textual' co-occurrences on the one hand or as 'semantic / syntactic' patterns on the other. However, it is possible not only to examine the internal syntagmatic properties of an expression, but also the pragmatic role of the expression in text and discourse. A third tendency therefore is to examine collocations in terms of performance, in other words from a **discoursal / rhetorical** point of view. From this perspective, idioms such as *to get the sack*, *to be fired* can be contrasted stylistically with less marked expressions: *to be dismissed*, *to lose one's job*. The difference between these expressions lies in their emphasis or rhetorical effect, as Moon (1987) and Fernando (1996) have argued. From a discourse analyst's perspective, Moon feels justified in arguing that syntactic and semantic constraints on fixed expressions are not as important as rhetorical function:

In general, studies of fixed expressions [...] concentrate on their typological and syntagmatic properties. Attention is given to such things as the degree of their lexical and syntactic frozenness, or their transformation potential: and even the primary characteristic of idioms, their non-compositionality as lexical units, may be seen as a matter of the interpretation of a syntagm. However, it is their paradigmatic properties which are of importance in relation to interaction. Fixed expressions represent meaningful choices on the part of the speaker / writer. (Moon 1994:117).

Fillmore and Atkins (1994) and Kay and Fillmore (1999) have similarly questioned the need for a distinction between idiom and collocation on the grounds of syntactic and semantic frozenness. Fillmore, Kay and O'Connor emphasised the fact that collocations are culturally salient items which need to be learnt as part of the language. According to their well-known definition, fixed expressions are:

[...] phenomena larger than words, which are like words in that they have to be learned separately as individual facts about pieces of the language, but which also have grammatical structure [and] interact in important ways with the rest of the language. (Fillmore, Kay and O'Connor 1988:501)



In a similar approach, Pawley and Syder have been influential in the area of language learning theory, and were among the first to emphasise that conversational gambits in natural speech were speech acts organised around fixed expressions of the type *it's easy to talk* (a reprimand for some criticism), *she's busy right now* (denying access by telephone) and *I thought you'd never ask* (expressing relief after permission has been granted) (1983:307). They pointed out that these expressions are effectively social institutions, and have specific cultural functions in the language:

A lexicalized sentence stem is a unit of clause length or longer whose grammatical form or lexical context is wholly or largely fixed; its elements form a standard label for a culturally recognized concept, a term in the language. Although lexicalized in this sense, most such units are not true idioms but rather are regular form-meaning pairings. (Pawley and Syder 1983:191-192).

This theme was similarly examined by Yorio, whose analysis of a spoken corpus found few traditional idioms, but instead proposed that sentence stems are key to understanding conventionalised fluency in language. Yorio concludes that grammatical accuracy must be matched by a knowledge of such idiomatic expressions:

Idiomaticity, or native-like quality in written language, appears to be a property characterized primarily by the presence of collocations and / or sentence stems rather than by actual idioms. [...] [A]lthough fluency is possible without grammatical accuracy, idiomaticity is not. Idiomaticity then becomes an excellent indicator of bilingual system proficiency and, as such, it deserves to be further studied and understood. (Yorio 1989:68)

Nattinger and DeCarrico (1992) examined shorter stretches of language than the sentence stem, and related knowledge of phraseology to a system of rhetorical expressions (1992:22). Following Coulmas (1979), they situated collocations within a continuum of increasing rhetorical force: from low to high impact. Nattinger and DeCarrico identified collocations as unmarked choices of expression '[co-occurring lexical items] that have not been assigned particular pragmatic functions by pragmatic competence' (1992:36). This 'unmarked' sense of the term collocation is an interesting departure from the perspectives we have seen above and clearly delimits the syntagmatic definition of collocation from a discursal one. Nattinger and DeCarrico then contrast unmarked collocations with lexical phrases, defined

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

as ‘marked’ collocations, in that they have recognised pragmatic functions. Lexical phrases are split into two groups (1992:38-42):

- Lexical units which do not allow paradigmatic or syntagmatic reformulation: polywords: *for the most part, as it were* and institutionalised phrases *how are you? what, me worry?*
- Grammatical frameworks with both fixed and free features: short range phrasal constraints: *a NP [time] ago*, long range sentence builders: *I think (that) [proposition clause X], the ADJ-er [proposition clause X], the ADJ-er [proposition clause Y]*.

The lexical phrase is proposed as an addition to the traditional distinction between idiom and collocation, and emphasises textual function rather than internal form:

Lexical phrases are parts of language that often have clearly defined roles in guiding the overall discourse. In particular, they are the primary markers which signal the direction of discourse, whether spoken or written. When they serve as discourse devices, their function is to signal, for instance, whether the information to follow is *in contrast to*, *in addition to* or is *an example of* information that it to proceed. (Nattinger and DeCarrico 1992:60)

According to Winter’s (1977) theory of clause relations, information in discourse is frequently managed lexically. Nattinger and DeCarrico show that this operates at a phrasal level by the use of global topic markers (*let’s look at*), shifters (*OK, now*) and summarisers (*so then*), as well as at a local level by the use of exemplifiers (*how about X?*), relators (*it has to do with Y*), qualifiers (*the catch is that...*), asides (*where was I?*) and so on. Such expressions are typical of the spoken language, but we see below that science writing has developed a sophisticated system for similar functions (including asides and topic shifters), albeit with different linguistic expressions. While such features may not be statistically significant across the corpus, and therefore do not usually figure in corpus-based analyses of register, Nattinger and DeCarrico claim that such phraseology has a significant role to play in the rhetorical construction of the text. These claims are supported by related studies on the pragmatic function of idioms in texts (Popiel and McRae 1988, Luzon-Marco 1999)

The ‘discourse / rhetorical’ approach is not concerned with lexis and grammar as such. Instead, the suggestion is that collocations and idioms can be distinguished on the basis of a rhetorical or textual function (as argued by

Nattinger and DeCarrico) or pragmatic marking (as argued by Moon). We have seen above that most idioms - such as *sell like hot cakes* (to sell quickly) and *pull a fast one* (to deceive by stealth) - are more marked stylistically than their typical paraphrases, not just for emphasis, but often with very specific information and a limited context of possible use. Moon has suggested that many such idioms and metaphors are deliberately used in speech and writing to bring in shades of evaluation or judgement in comparison with their unmarked equivalents (thus *the trial progressed at a snail's pace* would signal subjective feeling more explicitly than *the trial progressed slowly*). But as Moon points out, these 'prototypical' idioms are rarely found in authentic texts. In practice, the most commonly recurring expressions are likely to be 'lexical phrases' or 'sentence stems' and it is worth noting that apart from Nattinger and DeCarrico's work, these have received much less attention from lexicologists.

A normal text rarely moves in a clear-cut way from unmarked to marked expression, with idioms and collocations visibly demarcated. It is more realistic to picture a text as a sequence of different types of discourse signal, and while most of these expressions are idiomatic in that they have specific rhetorical or pragmatic roles to play, they are not marked as such within the normal reading of the text. Thus while lexical phrases may appear to be idioms from a traditional lexicological point of view, in their normal context they are simply part of the accepted phraseology. When something is 'marked' or pragmatically unusual, we can assume that it stands out from the expected style. Indeed, a knowledge of the expected phraseology is central to being able to step out of it in order to create some supplementary rhetorical effect. For example, Pawley and Syder's sentence stems have very specific and sophisticated rhetorical functions in spoken English: they are natural candidates for the category of idiom. But it does not make sense to suggest that they are permanently marked expressions, especially when we consider that they are commonly used in normal spoken discourse.

To give another example, the British English greetings *How do you do?*, *How are you?* *How do?*, *How's it going?*, *How goes it?* *Wotcha!* etc. vary from unmarked to marked in different contexts. The native speaker knows the core items (depending on dialect) and knows implicitly their rhetorical value in the phraseological system. *How do you do?* is felt to be the standard prototypical form, but this does not mean that it is the unmarked, neutral choice used in the majority of circumstances. The corollary of this is that prototypical expressions do not correspond to typical expressions. In addition, a notion of what constitutes 'collocation' or 'idiom' may also depend on an appropriate register or style and part of the meaning of an

idiomatic phrase is its specific context of use in which it is deemed to be appropriate (a pragmatic dimension rather than a strictly textual one). Thus from a discourse perspective, idioms (as relatively marked expressions) and collocations (as relatively unmarked expressions) might not be fixed categories, but may be perceived differently in different contexts. Collocations can be said to have a less fixed pragmatic set of uses than idioms; while lexical phrases, with their specific rhetorical roles, occupy a position somewhere in-between. From this basic premise, we can postulate a shifting rhetorical continuum between the usual phraseology of collocation and other more unusual expressions (including original expressions which break with collocational convention or stylistically marked idioms belonging to another discourse).

Collocation emerges throughout this discussion as a powerful but also extremely diverse concept. As van der Wouden (1997) notes, the term collocation itself either refers to the abstract relationship between words or the expression as a whole. Nevertheless, it is clear that although there are differences in application and methodology, all of the approaches we have summarised above converge on an important and recognisable phenomenon, the 'familiar recurrent expression'. Instead of arguing the case for one specific viewpoint, I attempt to see each as compatible and relevant at different points in my analysis. Since the main purpose of this book is to analyse a large corpus of texts, I argue below that the 'statistical / textual' perspective is the most appropriate approach to be adopted in the first stages of corpus analysis. However, the 'semantic / syntactic' perspective brings to our analysis of collocation the important notion of the abstract relationship between words, and the idea that the expression exists as a meaningful unit of choice within the grammar. The 'discoursal / rhetorical' view equally informs us of the role that the expression has within a running text and reminds us to interpret the expression as part of a system of stylistic alternatives. Despite differences of methods, each approach leads us to reconsider the relationship between words within the collocational expression and to revise the traditional notion of phraseology.

I intend to use the term **phraseology** to refer specifically to the rhetorical or pragmatic use of an expression. The term then stands in contrast to Halliday's 'lexico-grammar' which refers strictly to the cline between lexis on the one hand and grammatical systems on the other (Halliday 1985). The term also contrasts with the notion of 'collocational continuum' in lexicology (Howarth 1998), which refers to collocations as they become less like phrases and more like words. The 'discoursal / rhetorical' approach claims that the

pragmatic value of a particular expression constitutes an important aspect of a theory of phraseology. However, few studies of idiom or collocation have taken this perspective, and even fewer have attempted to account for systems of phraseology in scientific texts. My assumption in the analysis below is that although my collocational expressions are originally derived from the corpus on a statistical basis, they can be also usefully described in terms of their textual, rhetorical or pragmatic function. Thus a lexico-grammatical analysis of a specific discourse can be supplemented by an analysis of phraseology.

A further issue at this point concerns the notion of grammatical item (a closed class or functional word) and lexical item (an open class or content word). In the corpus analysis below, I suggest that grammatical items are useful starting points for the analysis of longer stretches of collocation and phraseology. We have seen in the discussion above that grammatical items have usually been left out of collocational studies. Many studies of textual collocation such as Phillips (1985) or Smadja (1993) go further and eliminate 'stop-words', largely because grammatical items are too frequent in the corpus and are reasonably thought to 'collocate with anything'. There is also a similar tendency in lexicology, in which grammatical items are usually considered only as collocations of lexical items (as with prepositional and phrasal verbs). However, as mentioned above, important work by corpus linguists such as Hunston and Francis (1998) on the patterns of grammar, and Renouf and Sinclair (1991) on consistent grammatical features of collocation has shown that grammatical items are fundamental to a theory of phraseology. The 'discoursal / rhetorical' approach has also brought into focus many previously ignored combinations of grammatical items which function as recognisable expressions. For example, many of Nattinger and DeCarrico's lexical phrases contain, ironically, very few lexical items: *just because, to be at it, as is, that's it then, it's all over, he's out of it*. These expressions are considered to be lexicalised, although they function more like utterances than single lexical items. Following on from this perspective, the analysis I set out below focuses on grammatical items as the key elements in longer stretches of phraseology. In section III, I specifically address the role of collocation in specialised texts and set out more fully Halliday's concept of the lexico-grammar.

The notion that grammatical items are closed class words will serve as my basic rule-of-thumb in order to identify these items. However, I also wish to explore the possibility that high frequency items (such as auxiliary verbs *is* and *has*) play an important role in the formation of collocations and fixed expressions, and assume therefore that such high frequency items are for the purposes of my analysis 'grammatical'. This frequency-based approach to

Christopher Gledhill (2000). *Collocations in Science Writing*. Tübingen, Gunter Narr, 7-20

lexis is consistent with Sinclair's view, and allows for a more nuanced analysis of words which are often considered to be at the intersection between grammar and lexis.