

Chapter 2

A Review of Related Literature on Collocation: Firthianism and Beyond

2.1 Introduction

At a time when few linguists, other than lexicographers themselves, devoted much attention to the study of lexis, and outlines of linguistics often contained little reference to dictionaries or other methods in lexicology, J. R. Firth repeatedly stressed the importance of lexical studies in descriptive linguistics (Firth 1935/1957a). He did not accept the equation of ‘lexical’ with ‘semantics’ (Firth 1951/1957a), and he showed that it was both possible and useful to make formal statements about lexical items and relations. For this purpose Firth regarded the statement about collocation as the most fruitful approach. This is expounded mainly in Firth (1951/1957a, 1952/1968, 1957b/1968) and some other places. Firth’s seminal ideas on collocation have since been developed by, among others, Halliday (1966), Mitchell (1966, 1971/1975) and Sinclair (1966, 1987, 1988, 1991). They developed the Firthian notion of collocation from different perspectives and with different theoretical emphases. This chapter is intended to review Firth and his three followers on the problem of collocation. We choose them to review and exclude other scholars mainly out of three considerations: First, they represent the mainstream of theoretical and practical study of collocation. Second, all of them consider simultaneously collocation and other lexical matters of the syntagmatic kind such as idiomatic expressions. As we emphasized in the last chapter, not to distinguish collocation from idiomatic expression is the theoretical cornerstone of the present study. Third, Firthian Linguistics, Hallidayan Scale and Category Grammar, and Sinclairian corpus linguistics are either the background of or closely related to Systemic Functional Grammar, the theoretical framework of the present research. In contrast to the general introduction of the first chapter, the review of the present chapter will focus on the perspectives and contributions of each scholar.

2.2 J. R. Firth: ‘You Shall Know a Word by the Company It Keeps’

It is over 250 years since Alexander Cruden¹ identified in his celebrated Concordance the repeated co-occurrence of certain words such as *dry* and *ground* in the Bible (see Kennedy 1998: 14, 108). In the 1930s, the British linguist H. E. Palmer working as an English language teaching specialist in Japan, wrote a monograph on collocation (see Mitchell 1975: 134; Kennedy 1998: 108). This research led Palmer (1933: 7) to conclude that collocations ‘exceed by far the popular estimate of the number of simple words contained in our everyday vocabulary’, thus calling for a reconsideration of the nature of vocabulary. As we mentioned in the last chapter, it is generally acknowledged that it is Firth who first introduced into descriptive linguistics the idea of collocation as a source of meaning. According to Firth, one can know a word from the company it keeps (Firth 1935/1957a, 1957b/1968: 175). In this section, we will first review the concepts of context and contextualization, which are at the heart of Firthian linguistics. Then, we will look at the Firthian notion of ‘meaning by collocation’. This is followed by a criticism on Firth’s collocation in particular and his theory of meaning in general.

2.2.1 Contextualization in Firthian Linguistics

Langendoen (1968), who misunderstood Firth in a number of ways (Halliday 1985: 7; Henderson 1987: 60), was right in believing that ‘the single most important notion in Firth’s early writings is that of context’. The belief in the importance of ‘context’ is basic to all Firth’s thinking (Henderson 1987: 60). Contexts for Firth may be of many kinds—phonetic, social and grammatical and so on. The idea of phonetic context was what led him away from the rather rigid monosystemic phoneme theory of Daniel Jones² towards the concept of polysystemic nature of language, with which his name was associated. His first account of Tamil pronunciation was presented in orthodox phonemic terms. He noted that the phoneme

¹From the 18th century, lists and concordances of words used in the Bible were made in an attempt to show that various parts of the Bible were factually consistent with each other. Alexander Cruden, a London bookseller, proofreader, morals campaigner and prison reformer, born in Aberdeen in 1710, produced the most famous of these for the Authorized (King James) Version of the Bible. First published in 1736, Cruden’s Concordance was a monumental piece of laborious scholarship which went through 42 editions even before 1879. It included concordances not only for what the author considered to be the major content words in the Bible but also some function words such as *how*, *you*, *he*, *once*, *between* and certain collocations such as *how long*, *how many*, *how much the less*, *all the nations* (see Kennedy 1998: 13–14).

²J. R. Firth went to University College London in 1928 as senior lecturer in phonetics under Daniel Jones, from where he transferred, in 1938, to the School of Oriental and African Studies.

/k/ had at least eight clearly perceptible variants [k̠, g, ɕ, k, g, x, γ, h̥] according to their position in the word and the neighboring sounds, i.e., according to their phonetic context. This would seem to be a straightforward allophonic statement. Other schools of linguistics have interested themselves in phonetic context to this extent, but usually as a way of accounting for what were felt to be insignificant or non-pertinent differences in pronunciation: the true linguistic units that the linguist was concerned with were what was left when contextual variation had been described and so disposed of. For Firth, however, contextual variations are an important inherent feature of language, not to be shrugged off as redundant or non-significant. Contextual variants are linguistic signals in their own right, providing information about linguistic structures which is not necessarily to be regarded as less important than the information provided by orthodox phonemes.

Context of situation is 'a key concept' (Firth 1950/1957a: 181) of Firth's theory of meaning and also the most important aspect of his thought of contextualization. For Firth, speech is personal and social activity interacting with the other forces in a situation. The cardinal principle is that language must always be studied as a part of social process and social activity, and every utterance should be considered and understood within its 'context of situation'. He acknowledged that he had taken this expression from the anthropologist Malinowski, who suggested that language was not a 'mirror of reflected thought', but a 'mode of action' (1923/1957a: 296). Firth felt, however, that Malinowski's context of situation was 'a bit of the social process which can be considered apart', that it was 'an ordered series of events considered *in rebus*', whereas his own context of situation was an abstraction, 'a schematic construct', 'a group of related categories at a different level from grammatical categories, but rather the same abstract nature'. The categories, he suggested, were (1) the relevant features of participants: persons, personalities; (2) the relevant objects; (3) the effect of the verbal action. In addition to these interior relations of the context, he later suggested that there should be reference to more general frameworks such as economic, religious, and other social structures, types of discourses (monologue, narrative, etc.) (Firth 1957b/1968: 178). Firth even prophesied that 'sociological linguistics would be the great field for future research' (1935/1957a), although this has not been materialized up till now with the mainstream of modern linguistics focusing almost entirely upon the individual and his psyche. Firth's thought of contextualization is fully represented by the following remarks on what he called 'contextual elimination':

The moment a conversation is started, whatever is said is a determining condition for what, in any reasonable expectation, may follow. What you say raises the threshold against most of the language of your companion, and leaves only a limited opening for a certain likely range of responses. This sort of thing is an aspect of what I have called contextual elimination (Firth: 1935/1957a: 31–32).

The above is the background against which we consider the Firthian notion of collocation, the lexicogrammatical aspect of contextualization.

2.2.2 *Meaning by Collocation*

Firth (1951/1957a) introduced ‘collocation’ as one of ‘the lower modes of meaning’ and illustrated it by the claim that part of the meaning of *ass* in modern colloquial English could be stated in terms of collocation:

- (a) An ass like Bagson might easily do that.
- (b) He is an ass.
- (c) You silly ass!
- (d) Don’t be an ass.

One of the meanings of *ass* is its habitual collocation with an immediately preceding *you silly*, and other phrases of address or of personal reference. Even if you said ‘An ass has been frightfully mauled at the zoo’, a possible retort would be, ‘What on earth was he doing?’ There are limited possibilities of collocation, the commonest adjectives being *silly*, *obstinate*, *stupid*, *awful*, and occasionally *egregious*. *Young* is much more frequently found than *old*. The plural form is not very common. Firth (1951/1957a: 195) pointed out that ‘meaning by collocation is not at all the same thing as contextual meaning, which is the functional relation of the sentence to the processes of a context of situation in the context of culture’. ‘Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night*’ (Firth 1951/1957a: 196). This was his alternative to the referentially based theories of word meaning passed over above (Martin 1992: 275).

Firth (1951a/1957: 195) classified the study of distribution of common words into ‘general or usual collocations’. On the one hand, the commonest sentences in which the words *horse*, *cow*, *pig*, *swine*, *dog* are used with adjectives in nominal phrases, and also with verbs in the simple present, indicate characteristic distributions in collocability which may be regarded as a level of meaning in describing the English of any particular social group or indeed of one person. This is what Firth called ‘general or usual collocations’. On the other hand, the study of the usual collocations of a particular literary form or genre or of particular author makes possible a clearly defined and precisely stated contribution to what Firth termed the ‘spectrum of descriptive linguistics’, which handles and states meaning by dispersing it in a range of techniques working at a series of levels (Firth 1951/1957a: 192). This is what Firth called ‘technical or personal collocations’. Firth further illustrated collocations of this kind from Lear’s limericks, the poems of Swinburne,³ and letters of the eighteenth and early nineteenth century. All of them

³Swinburne, Algernon Charles (1837–1909). English poet and critic who introduced prosodic innovations and who became noteworthy as the symbol of mid-Victorian poetic revolt. He was also an important and prolific literary critic of the later 19th century. [K. Kuiper (ed.), 1995, Merriam Webster’s Encyclopedia of Literature, Springfield, Massachusetts: Merriam-Webster, Incorporated, Publishers, p. 1084].

belong to 'restricted languages', which may be said to have their own 'micro-grammar' and 'microglossary' (Henderson 1987: 62).

2.2.3 *Criticism of the Firthian View of Collocation*

It would be wrong to claim that Firth's view of collocation has secured uncritical acceptance among a majority of linguists. Lyons, in his paper *Firth's theory of meaning* (Lyons 1966: 288–302), attacks Firth's view of meaning in general and collocation in particular:

Although he devotes considerable space to a discussion of his concept and constantly refers to the 'collocational level' as intermediate between the grammatical and the situational, he never makes clear how the notion of collocation fits into his general theory. His assertion that 'one of the meanings of *night* is its collocability with *dark* and of *dark*, of course, collocation with *night*', would seem to bring 'the statement of meaning by collocation' in line with the distributional theory of meaning advocated by Harris and Hoenigswald. And the distributional theory of meaning is very quickly disposed of on at least three counts: firstly, it does not satisfy the conditions of material adequacy governing the use of the term of 'meaning'; secondly, it appears to involve the identification of language and text (or of 'langue' and 'parole'); and finally, even if it were true that similarity and difference of distribution could be correlated with similarity and difference of meaning, there are many other more important meaning relations...and these relations cannot be derived by purely distributional, or collocational criteria... (Lyons 1966: 295).

Lyons remarks that Firth 'never makes clear how the notion of collocation fits into his general theory'. This is surely not true: the lexical level is one of the 'modes' into which meaning is dispersed, and some lexical items co-occur with a high degree of regularity, thus providing syntactic contexts for each other. Since meaning is interpreted as function in context, the function of a lexical item in a collocation is, for Firth, part of the meaning of that item. As for the problems with distributional theory of meanings of Harris and Hoenigswald, we do not think they are relevant here. Firth's purpose of using distributional criterion for collocation study is only that it is much more scientific than the traditional referential approach to language teaching. Lyons points out that Firth's theory has no place for 'sense relations' of the type characterized by later structural semanticists. Again it is important to take account of the overall aims and orientation of Firth's theory. Just as Butler (1985: 11) argues, Firth would have been skeptical of much of the later work on structural semantics, feeling that it forced language into the straitjackets of neat oppositions. Take, for instance, the sense relation of synonymy. As is well known, very rarely, if ever, are two or more lexical items fully intersubstitutable in all contexts. Given Firth's view of meaning as function in context, this leads inexorably to the view that there are no true synonyms. Firth would certainly not have wanted to elevate 'cognitive meaning' above other types of meaning, so that for him even the 'cognitive synonymy' would be unacceptable.

Firth's concept of meaning by collocation has also been attacked by Langedoen. Langedoen (1968) rejects Firth's 'meaning by collocation', claiming that 'the goal

of semantics should be, rather, to show how meaning of such phrases as *dark night* is determinable from a knowledge of the meaning of the lexical item comprising them and the syntactic relationships that are found in them' (Langendoen 1968: 64). But here Langendoen has failed to appreciate the significance of Firth's pronunciation that collocation is merely an abstraction from the syntagmatic patterning of lexical items, and is not to be confused with other types of meaning, including the 'semantic meaning' (a perfectly proper and consistent term for Firth) which Langendoen regards as primary. Again, as with Lyons, there is a confusion of aims: Langendoen appears to assume that the only kind of phrasal meaning which is to be accounted for is the 'compositional' meaning derived from the presumably context-independent paradigmatic sense relations of the individual items; Firth, on the other hand, while not rejecting compositionality, is concerned to account for the function of each linguistic item in its context, and points out, correctly and relevantly, that part of this context is the linguistic environment provided by other lexical items which are co-present. Langendoen may feel that Firth's goals are not worth pursuing (though he does not give convincing reasons why they are not); but it should be recognized that if these goals are accepted, Firth was being entirely consistent in his treatment of collocational meaning.

Langendoen (1967: 62) also argued that Firth's proposal for studying words in lexical contexts amounts not to assertions about semantics but about stylistics. It is true that there is a stylistic dimension in Firth's theory of collocation. Firth (1951/1957a: 190) admits that his paper is an attempt to 'show the dangers of an over-facile superficial use of the word stylistics, without an adequate logical syntax or even without referring to the higher levels of the spectrum of meaning, such as those provided by the biographical and cultural contexts' (1951/1957a: 203). He argued that the statement of Swinburne's philosophy by a philosopher would be impossible without a previous analysis of his language. But it is only part of the story. Firth (1951/1957a: 190) also made it clear that the general purpose of the study is to 'sketch the framework of a language of description in English about English for those who use English, to illustrate what I understand by linguistic analysis'. Obviously, stylistics is not the only purpose of Firth's paper. Furthermore, Langendoen's claim loses sight of the fact that Firth did distinguish between 'general/usual collocations' and 'technical/personal collocations'. The large amount of linguistic description of collocation by Firth's followers has also proved Langendoen's claim unreasonable.

In our opinion, Firth's collocation should be understood against the background of his interpretation of meaning. The patterned arrangement of linguistic form, variously phonological, lexical, and grammatical, is its own strand of meaning separable from all other types of meaning yet intertwined with them. The formal value of an item depends closely on other items present in the text and the constraints and dependencies observable between them. Collocation is the abstract arrangements of a lexical or lexicogrammatical kind. Firth insisted that it was important to consider syntagmatic aspects of a lexical item's meaning potential. Firth's mutual expectancy between lexical items was one aspect of Firth's interest in the notion of predictability in discourse. What Firth put high value on is the

dependency between sentences and in particular between elements of a sentence, the syntagmatic relations obtaining between parts of discourse. Even if Firth's theory was deficient with regard to sense relations of a paradigmatic kind, it can be claimed that it paid more attention to the syntagmatic aspects of sense relations, as represented by collocation, than have most semanticists.

2.3 M. A. K. Halliday: Lexis as a Linguistic Level

According to our investigation, the notion 'the level of collocation' first appeared in Firth (1952/1968: 16). In Firth (1957b/1968), 'collocational level' appeared in contrast to 'context' and 'citation':

...it will be seen that collocation is not to be interpreted as context, by which the whole conceptual meaning is implied. Nor is it to be confused with citation... Lexicographical citations are keyed to the definitions intended to exemplify a series of different meanings arbitrarily selected and defined, and also to illustrate changes of meaning. The habitual collocations in which words under study appear are quite simply the mere word accompaniment, the other word-material in which they are most commonly or most characteristically embedded. It can be safely stated that part of the 'meaning' of *cows* can be indicated by such collocations as *They are milking the cows*, *Cows give milk*. The words *tigresses* or *lionesses* are not so collocated and are already clearly separated in meaning at the **collocational level** (Firth 1957b/1968: 180, emphasis by this author).

It seems that Firth introduced 'collocational level' mainly as one intermediate between the grammatical and the situational level. As for how much lexis is independent of grammar, Firth himself seemed to have no opinions. It is his follower, notably Halliday, who took the matter to the point of regarding lexical study as independent of grammar. Halliday (1966) uses 'lexical level' in preference to 'collocational level' in order to suggest greater generality and parallelism with grammatical level. He suggests that lexis may be usefully thought of (1) as within linguistic form, and thus standing in the same relation to (lexical) semantics as does grammar to (grammatical) semantics; and (2) as not within grammar, lexical patterns thus being treated as different in kind, and not merely in delicacy, from grammatical patterns (Halliday 1966: 148). Halliday's (1966) idea of 'lexis as a linguistic level' was expressed as early as in 1964 (Halliday et al. 1964) and it echoed in Sinclair (1966). What this idea highlights is the differences between lexis and grammar as two ways of looking at the internal pattern of language, which can be summarized as follows: (1) Grammar involves 'system' type choices: those where there is a restricted set of contrastive possibilities. *this/that* and *who/whose/what/which* each form a grammatical system in which the items themselves are terms; these are 'fully grammatical' items and we can account for the whole of the contrast between them in grammar. *this*, for example, can be defined as not *that*. Lexis, on the other hand, is 'class' type patterning in linguistic form. We cannot, in grammar, distinguish between items *table*, *desk* and *bench*. *table* does not equal *not desk* or *not desk or bench*. There is no system in which *table* is a term. *table*, *desk*

and *bench* are members of a lexical set. (2) Grammatical systems are bounded only by probability. Whereas in grammar we can say: ‘at this place in structure, these terms are possible, and all others are impossible’. Lexical sets, on the other hand, are bounded only by probability. Given the items *chair*, we are more likely to find in the same utterance the item *sit* or *comfortable* or *high* than, say, *haddock* or *reap*, though no one could maintain that the latter are impossible. (3) While grammatical patternings are organized in rank, i.e., word, group/phrase, clause. Lexical sets are not organized in rank. A lexical item, as a formally prepackaged bundle of meaning, can range over the whole rank scale, from word (e.g., *grammar*) to group/phrase (e.g., *to and fro, it’s not in his nature to...*) to clause (e.g., *it’s raining cats and dogs*). (4) A lexical item may not be co-extensive with grammatical item. In lexis, we are not interested in the difference between *take*, *took* and *taking*, which is purely grammatical and does not affect collocation; but we are interested in the fact that *take off* collocates with different items from those that *take over* collocates with, and both differ in this respect from *take*. Halliday et al. (1964) showed this in Table 2.1.

In grammar we distinguish four items: (1) a word *took*, (2) a word *taking*, (3) a morpheme *take* and (4) a word *take*; but these are the same items whether followed by *off* or by *over*. In lexis on the other hand we distinguish two items: (1) *take off* and (2) *take over*; but *take off*, *taking off* and *took off* are all the same item.

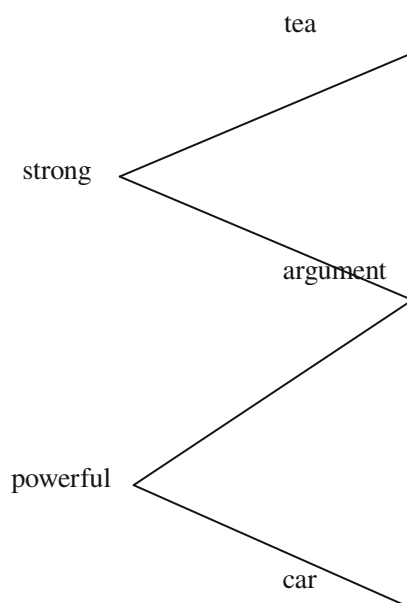
If Firth (1951/1957a) had already, to a certain degree, separated lexical matters from semantics and grammar, Halliday (1966) was concerned to make that separation more complete. The many unresolved issues of language patterning left over when grammatical analysis, however thorough, was complete, could either be relegated to semantics or tackled at a lexical level of analysis, with the aim of making lexical statements at the greater level of generality than dictionaries do. As an example of the lexuality of collocation, Halliday (1966) compares the different collocability of *strong* and *powerful*. The figure below shows the acceptability of *strong tea* but not of *strong car*, while argument collocates with both. Moreover, the

Table 2.1 The lack of co-extensiveness of lexis and grammar (from Halliday et al. 1964: 36)

	Lexis item (1)	item (2)
Grammar		
item (1)	the aircraft <u>took off</u>	the captain <u>took over</u>
item (2)	<u>taking</u> off was easy	<u>taking</u> over was easy
item (3)	a neat <u>take</u> off	a neat <u>take</u> over
item (4)	about to <u>take</u> off	about to <u>take</u> over

Note Lexical items are indicated by a line above and grammatical items by a line below

Fig. 2.1 The lexicality of collocation



relation is constant over a variety of grammatical configurations: *He argued strongly against...*; *the strength of his argument*; *This car has more power*; etc. So the lexical statement can operate independently of grammatical restrictions. *Strong*, *strength*, *strongly*, *strengthen* represent the ‘scatter’ of the same lexical item (Fig. 2.1).

Generally speaking, what Halliday (1961, 1966) was doing is to highlight the difference between lexis and grammar as two types of language patterning, and to introduce a category of ‘lexicalness’ to parallel that of ‘grammaticalness’ under his theoretical framework of Scale and Category Grammar, or the so-called New-Firthian Linguistics. At this period of Hallidayan theory, it seems better to treat lexical relations as on a different level, and to require a different theory to account for them, linguistic forms which grammar cannot handle. As the shift in the center of gravity of the model from form to semantics in the 1970s, the rapprochement of grammar and lexis was favored by developments in the systemic model of Hallidayan theory and the difference between lexis and grammar was subsumed by the concept of ‘lexis as most delicate grammar’, which we will discuss in the next chapter.

2.4 T. F. Mitchell: Linguistic Goings-on

Another name central to the British approach to lexis is F. Mitchell. He was one of the younger colleagues of Firth and made his name by the seminal work of applying Firth’s framework of context of situation to the language of buying and selling in

Cyrenaica (Mitchell 1957). Mitchell's work represented one of the earliest attempts to study genre in the British tradition of linguistics and 'it was not followed up within systemic linguistics for almost a generation' (Martin 1992: 504). His interest in lexis can be noticed from his paper devoted to the paper collection in memory of J. R. Firth (Bazell et al. 1966), which is entitled *Some English Phrasal Types*. This paper is intended primarily to draw attention to what is felt by the author to be the need for greater explicitness in the definition of linguistic items and their classes. In another paper of his, *Linguistic 'goings-on': collocations and other lexical matters arising on the syntagmatic record*, Mitchell (1971/1975) made an attempt to develop the Firthian concept of collocation and also considered interrelated lexical categories of a syntagmatic kind such as compound and idiom. Mitchell's concern is principally with 'forms of language', but this interest is located against the background of a total approach to meaning. The following is a review of the main points of this paper.

2.4.1 *The Formal Meaning of Language*

Following the Firthian Linguistic tradition of the study of meaning, Mitchell (1971/1975) does not feel satisfied with the common enough view of semantics as concerning either the meaning of lexical items taken singly or that of a text over and above the meaning of its lexical items. His paper is intended to show that 'it is as unsatisfactory to regard "word-meanings" as discrete, finite, and localized in advance as it is to see "further" meaning as a kind of "expressive" topping or dressing (ibid.: 99). For him, language—and meaning with it—is not only contrastive but also on-going. Dependency between sentences and in particular between elements of a sentence, otherwise the syntagmatic relations obtaining between parts of discourse, is Mitchell's recurring theme in this paper. In order to illustrate this character of language, Mitchell uses two idiomatic expressions, *That'll do* and *silly old bastard*, as example:

This expression, like an infinite number of other sentences in English, is ambiguous in isolation and only disambiguated in extended discourse. With final rising intonation, it may be used homophonously for example by dad either to his erring small boy or to the shop assistant looking him out a tie, but a repetition of the form of words, this time with falling intonation, is adapted to the continued recalcitrance of the child but hardly to the requirements of the shop transaction. Another example is the addressive English sentence *silly old bastard*. This expression is usually used to indicate the speaker's appreciation of a favor rendered by a close friend and it impugns neither the addressee's intelligence, his vigor, nor his antecedents. The meliorative employment of the sequence cannot be derived from the amalgamation of the meaning of such 'underlying' sentences as *he is silly, he is old, he is bastard*. From these examples comes the recognition that meaning not only resides in contrastive relationships but also in the on-going nature of language (Mitchell 1975: 101).

Mitchell (1971/1975) at first distinguishes seven categories of meaning: functional meaning, emotive meaning, topical meaning, socio-cultural meaning,

ostensive meaning, referential meaning and mnemonic meaning. They are seven basic types of use that language is put to. The web of meaning is made up of many strands and more than one of the seven types of meaning may meet at one node of the web, at a single place in discourse. The seven types of meaning constitute the background against which Mitchell evaluates the forms of language. The formal meaning of language is ‘of paramount importance for linguists’ (Mitchell 1971/1975: 102). The patterned arrangements of linguistic form, variously phonological, lexical, and grammatical, is its own strand of meaning separable from all others yet intertwined with them. Mitchell’s paper is concerned with abstract arrangements of a lexicogrammatical kind. The formal value of an item depends closely on (1) other items present in the text and the constraints and dependencies observable between them, (2) the ‘permutability’ of the text in terms of the analytical operations of substitution, expansion or contraction as the case may be, interpolation (a form of expansion), and transposition. (1) may be termed intra-textual dependence and (2) inter-textual dependence. (1) is almost certainly consequent upon (2). Take the particle *off* as an example. The presence of *off* in *the milk has gone off* (expansion), according to Mitchell, marks a different *gone* from that in *the milk has gone* (contraction), just as a different *off* is almost certainly ‘determined’ by the presence (substitution) of *John* in place of *the milk* (*John has gone*). Many other features of distinctive linguistic behavior will relate to (*go* $\pm$ *off*) and, in turn, the lexicographer-grammarian must go on to consider (*go* + *off*) and to distinguish at least between *go off* (=away), *go off* ($\pm$ the field), *go off* (someone), *go off* ($\pm$ bang), *go off* (one’s head), (milk) *go off*. Mitchell argues that the formal linguistic patterning, i.e., the intra- and intertextual constraint and dependency of language items, should be the top concern of the study of meaning.

2.4.2 Collocation as Association of Roots

It is taken for granted that collocation is the association of ‘words’. Mitchell, however, points out insightfully that ‘linguistic analysis has been bedeviled by the use of ‘word’ as both several kinds of linguistic units and as the names for those units’ (Mitchell 1971/1975: 120). For him, the concept of ‘collocation’ has to be seen partly in relation to that of ‘root’ or ‘lexeme’. A recognizably regular association of roots or a collocation undergoes flectional variation which is accountable in terms of the extended context. For example, the last two words of *he drinks heavily* contains marks of verbal agreement with the preceding third person singular subject (-s in *drinks*) and of an ‘adverbial adjunct’ (-ly in *heavily*); comparable word forms are *heavy* and *drinker* in *he is a heavy drinker*, *heavy* and *drinking* in *he is putting in some heavy drinking*, and *drinking* and *heavily* in *he is drinking pretty heavily*. Collocations, therefore, are of roots ($\sqrt{\text{heav-}}$ and $\sqrt{\text{drink}}$ in the above case), not of words, which are essentially means of reference (Mitchell 1971/1975: 117). For most practical purposes, the root may be identified as the HCF of *relata* within a scatter, so that odd-looking roots like $\sqrt{\text{edu-}}$ and $\sqrt{\text{polit-}}$ are recognizable as

contributing to collocations ($\sqrt{\text{edu-}} \sim \sqrt{\text{system}}$) ($\sim$ = 'transposable with') in *an educational system* and *a system of education* and ($\sqrt{\text{party-}}$ and $\sqrt{\text{polit-}}$) in *political party*, *party politics*, *party politician*, and *party political broadcast*.

Collocational analysis, for Mitchell, has at least two important objects. The first object is to provide palpable identity for abstract roots, whose putative central cores or features are forever so maddeningly elusive and which are so ill-defined by the application to them of vague aprioristic notions and glosses deriving in any case, however unconsciously, from the use of root *in extenso*. To arrive at the meaning of any element of linguistic structure, it first behoves us to put it back where it came from and, unlike linguists, human beings do not speak in roots. To take the adjectival paradigm only of the root $\sqrt{\text{heav-}}$, there is clearly no other lexical item in English regularly associated *inter alia* with the roots of *cold/blow/dew/soil/damage/sarcasm/sky/drinking/breathing/make-up/hand/crop/rain/work/lorry/gun/accident/fall/heart/features/top-spin/humor/hydrogen/meal/going/etc.* Roots themselves, however, are zero collocations. The second purpose of collocational study is to recognize the root + elements which discourse further comprises. A collocation is a 'composite structural element' in its own right. If the reader takes the trouble to sift the foregoing collocations containing adjectival $\sqrt{\text{heav-}}$ through their several grammatical distributions, he will find very little matching from one collocation to the next. Mitchell (1971/1975: 119) gives two examples to illustrate this point: ($\sqrt{\text{heav-}} \sim \sqrt{\text{drink-}}$) and ($\sqrt{\text{heav-}} \sim \sqrt{\text{damage-}}$). The first occurs in the grammatical patterns Adjective + Agentive Noun (*heavy drinker*), Verb (intransitive) + Adverb (=Adj + -ly) (*to drink heavily*), Adj + Gerund (*heavy drinking*), and in the compound adjectival form involving adjectival -ing [(a) *heavy drinking (man)*]. The second example, on the other hand, is distributed among Adj + Non-agentive Noun (singular) (*heavy damage*), Verb (transitive) + Adverb (=Adj + -ly) (*damage heavily*), Adv + Passive Particle (*heavily damaged*). Any matching is clearly minimal. Mitchell concludes that underlying all such present distinctions and those recognized in the subsequently extended frame of reference is the notion of contrast, by which a linguistic item or class of items derives its meaning from the place it occupies and the contrasts it develops within widely ramifying networks of differential relations.

2.4.3 Colligation, Idiom and Compound

Other lexicogrammatical 'composite elements' considered by Mitchell (1971/1975) are colligation, idiom and compound. Colligation, a term coined by J. R. Firth, is used by Mitchell to refer to the generalizable classes of collocations. As collocations are nameable by words, so colligations involve the use of word-classes to name the collocational class. Colligational labels underline the necessary admixture of 'functional' and 'formal' as in the case of ('motive' verb + 'directional' particle, describing *tear/lope/race* etc....*up/along/across* etc.). Mitchell (1971/1975: 121) argues that as in the case of the individual collocations making up the class, the

colligation is to be seen as an entity, therefore a better labeling for the above case should be ['motive' (verb + particle)]. The relationship between 'collocation' and 'colligation' is therefore on the scale of generality. For Firth, colligation is the interrelation of categories used to make the statement of meaning at the grammatical level, while collocation is the mutual expectancy of lexical items (Firth 1957b/1968). Mitchell, however does not agree to separate artificially syntax from lexis, which are both covered by the syntagmatic relations obtaining between linguistic items in texts.

Mitchell (1971/1975) distinguishes between collocations and idioms. An idiom resembles rather a root; it is a bloc or assemblage of roots, non-productive in terms of the productivity of roots within it. It is a particular cumulate association, as a rule inoperable in the sense that its parts are unproductive in relation to the whole in terms of the normal operational processes, that of substitution above all. The collocation *tear up* is not an idiom because there is no such fixity of association between *tear* and *up*. *lope*, *amble*, *shamble*, *race*, etc. may be substituted for *tear*, and *down*, *across*, *onto*, *into*, *along*, etc. for *up*. An idiom is an entity whose meaning cannot be deduced from its parts. For example, *tear down* in *he put down the book* is a collocation while *put down* in *he put down the rebellion* constitutes an idiom. Certainly, Mitchell also notices the connection or similarity between collocation and idiom and the difficulty of making a sharp distinction between the two. First, idioms can occur as part of collocations (e.g., [*the nose on your face*] in *as plain as [the nose on your face]*) or combine to form a collocation (e.g., [*take off*] (= imitate) + [*to a T*] (=perfectly) in [*to take (someone) off to a T*]). Second, both idioms and collocations often correspond to a cognitively similar single form which may replace them either optionally or obligatorily in certain (stylistic) contexts: idioms, *make up* = *compose*, *make it up* = (be) *reconcile(d)*, *make up to* = *flatter*; collocations, *put down (the book)* = *deposit*, *come down* = *descend*. Third, Collocations and idioms are similar to the extent that both are generally relatable to grammatical generalizations and that both cut across syntactic classes [e.g., verb + object complement: *play tricks* (collocation), *kick the bucket* (idiom), verb + adverbial complement, *put on (the coat)*(collocation), *put off (the meeting)* (idiom)]. Fourth, the myriad functional uses to which language is put day by day (functional idioms such as *do you think you could...*, *could you possibly...*) may confidently be expected to correspond to the extensive appearance of idioms and collocations. Mitchell (1971/1975: 128) points out all these 'pieces' of language may derive from the fact that we make use of the comparatively limited number of morphemes and morphemic patterns, of phonatory potentialities, and indeed of generalized formal patterns of grammar, even of sentential type, for the manifold functions of language. Another kind 'composite element' considered by Mitchell in relation to collocation is compound such as *bullfight*, *high chair*, *billiard ball supply company*, etc. The individual compound is essentially uninterruptable. Compounding is a process of 'coalescence'. It is not an additive one, as in the case of attributive adjectives of color [e.g., *white* in *white (billard balls)*], which permit

‘modifications’ of other kinds, e.g., adjectival comparison (*white/-er/-est*) or ‘adverbial’ intensification (*very white*, etc.). The three kinds of ‘composite elements’ considered by Mitchell are not distinguished strictly in the present research for reasons we illustrated in detail in the last chapter (see Chap. 1 of this dissertation).

To summarize, Mitchell (1971/1975) develops the Firthian notion of collocation in many ways. He emphasizes the importance of considering formal meaning of language in semantics. He highlights the on-going nature of language. He is essentially concerned with all kinds of syntagmatic delimitation and his work represents a unique blend of levels of analysis, a syntagmatic-lexical approach similar to that of Sinclair in the COBUILD project. His paper, of which the merits have been far from fully recognized and appreciated, I believe, deserves the detailed review above.

2.5 J. Sinclair: Corpus Linguistic Approach to Collocation

The largest practical collocation studies so far reported is that of Sinclair and his colleagues’ (Sinclair 1966, 1987, 1988, 1991; Sinclair et al. 1991). This research has taken full advantage of the development of the computer technology in the second half of the last century and is known as the corpus linguistic approach to collocation: a way of investigating language by observing large amount of naturally-occurring, electronically-stored discourse, using software which selects, sorts, matches, counts and calculates. The initial stages of this research were made possible by a grant from the Ford Foundation to the University of Edinburgh and later by a grant from the office for Scientific and Technical Information. The result was reported in the paper *Beginning the Study of Lexis* (Sinclair 1966). In this paper, Sinclair identified theoretical and descriptive issues and developed ideas which were not brought to fruition until the 1990’s (Carter 1991). During the 1970s, computational research on English did not progress much in Birmingham because all the energy was spent on preparing for the future—devising software packages, instituting undergraduate courses, and influencing opinions on the campus. At the beginning of 1980s, the English Language Research at the University of Birmingham teamed up with Collins publishers to investigate lexis and produce reference books. The essential preparatory work done in the 1970s was therefore put to good effect, and transformed through experience into a completely new set of techniques for language observation and analysis. Theoretical findings are reported on various occasions (Sinclair 1987, 1988; Sinclair et al. 1991) and largely appear in the book *Corpus Concordance Collocation* (Sinclair 1991). This book ‘sets out an agenda for a radical departure in the description of English’ (Hunston and Francis 2000: 14) and ‘represents both a culmination and a new beginning’ (Carter 1991). In the following discussions we will review some of the fundamental aspects of Sinclair’s theoretical findings.

2.5.1 *The Idiom Principle*

Sinclair (1991: 109) contends that in order to explain the way in which meaning arises from language text, we have to advance two different principles of interpretation: the open-choice principle and the idiom principle. The open-choice principle is a way of seeing language texts as the result of a very large number of complex choices. At each point where a unit is completed (a word, phrase, or clause), a large range of choice opens up and the only restraint is grammaticalness. This is probably the normal way of seeing and describing language. It is often called a 'slot-and-filler' model, envisaging texts as a series of slots which have to be filled from a lexicon which satisfies local restraints. This model is clearly illustrated by the Chomskyan approach to grammar: the nodes on the tree structure are the choice points. According to Sinclair (1991: 110), 'virtually all grammars are constructed on the open-choice principle'.

The open-choice principle does not provide for substantial enough restraints on consecutive choices. We would not produce normal text simply by operating the open-choice principle because words obviously do not occur at random in a text. Meaning of texts are made in chunks of language that are more-or-less predictable, though not fixed, sequences of morphemes lead Sinclair to an articulation of the 'idiom principle':

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 analyzable into segments. To some extent, this may reflect the recurrence of similar situations in human affairs; it may be motivated in part by the exigencies of real time conversation. However it arises, it has been relegated to inferior position in most current linguistics, because it does not fit the open-choice model. (Sinclair 1991: 110, emphasis by this author)

The idiom principle can be seen in the apparently simultaneous choice of two words, for example, *of course*. This phrase, according to Sinclair, operates effectively as a single word, and the word space, which is structurally bogus, may disappear in time, as we see in *maybe*, *anyway*, and *another*. Just as it is misleading and unrevealing to subject *of course* to grammatical analysis, it is also unhelpful to attempt to analyze grammatically any portion of text which appears to be constructed on the idiom principle (Sinclair 1991: 113). The two fundamental realization forms of the idiom principle are collocation and idiom (Sinclair 1991: 173), which are both covered under the term 'collocation' in this research.

Both the idiom principle and the open-choice principle are described as 'ways of seeing or interpreting language'. In other words, a language user, faced with an instance of language use, has to decide whether to interpret this as a chunk, or as a series of individual items. Take, the word sequence *I must confess*, for example. The meaning of this sequence may be paraphrased as 'I am going to tell you something you may find unpleasant, or something I find embarrassing' (as in *I must confess I rather like Alice*). There is evidence that the sequence *I must confess* acts as a single item here, in that the various parts are not freely substituted with other

words. *He must confess*, or *I must not confess* are not frequently found in this sense. If, on the other hand, the sequence *I must confess* is interpreted according to the open-choice principle, the meaning may be paraphrased as ‘I am under an obligation (possibly self-imposed) to admit to a wrong-doing’ (as in *The police have found my finger-prints on the gun. I must confess*). In this case, other words may be substituted (*I must confess; I must not confess; I must run away*). The point is that the hearer, on hearing *I must confess* in any circumstances, must decide which meaning is appropriate. Both interpretations are not simultaneously possible. For normal texts, Sinclair (1991: 114) proposed that the first mode to be applied is the idiom principle, since most of the text will be interpretable by this principle. Whenever there is good reason, the interpretive process switches to the open-choice principle, and quickly back again. In this sense, Sinclair seems to give some priority to the idiom principle of interpreting language.

2.5.2 *Sense and Structure: The Invasion of Lexical Hordes*

Sinclair’s work does not stop with the observation of collocation, or with the compilation of dictionaries. Sinclair is also ‘a grammarian of distinction’ (Tucker 1996: 147). He approaches grammar through the words, and the word as observed in the corpus. His approach can be basically described as to gradually reveal the collocational and grammatical behavior of words by analyzing their environments with the help of computer. To uncover the grammar of a language then, in a Sinclairian manner, is to construct it out of the dictionary, to build it step by step from the behavior of individual words. This approach to grammar is fully demonstrated in Sinclair (1987, 1991). He uses the term ‘structure’ to indicate a lexical item and its patterns and collocations. He defines it as ‘any privileges of occurrence of morphemes’, whether those morphemes are lexical (as in the collocation of *yield* and *profit*), or grammatical (as in the collocation of *yield* and *up*) (Sinclair 1991: 104). ‘Structure’ therefore encompasses both we refer to as ‘pattern’ and further collocational regularities. Chief among the new insights gained by Sinclair is the realization of the delicate relationship between sense and structure, that is, the different senses of an item are often paralleled by preferred structural configurations. Sinclair makes the point about sense and structure largely in the context of distinguishing between senses of polysemous words. If a word has several senses, and is used in several patterns, each pattern will occur more frequently with one of the senses than the others, such that the patterning of an individual example will indicate the most likely sense of the word in that example. This is demonstrated with respect to three words: *decline*, *yield* and *set* (Sinclair 1987). Sinclair (1987: 109–110) summarizes his argument by referring to his experience when working with compilers of the 1987 COBUILD dictionary, when compilers were asked to identify the most typical examples of a (sense of a) word and to describe their structure:

In nearly every case, a structural pattern seemed to be associated with a sense...In the vast majority of cases, the compiler, in choosing typical instances, had little doubt about the kind of syntactic pattern that would have to be featured. And more briefly: 'It seems that there is a strong tendency for sense and syntax to be associated' (Sinclair 1991: 65).

Sinclair prophesizes that if this can be demonstrated by further research to be a central and not an occasional feature of language, then it can be turned to provide valuable evidence for lexicography—suggesting sense division, and phraseology—identifying phrase units with distinctive patterning. 'The traditional domain of syntax would be invaded by lexical hordes' (Sinclair 1987: 87).

2.5.3 *Units of Meaning*

The observation of the idiom principle and the association between sense and structure raises issues concerning units of meaning of language description. When discussing the relationship between sense and structure, Sinclair comments that, in particular with relation to recurring phraseologies associated with common words, we find 'a distribution of meaning across a number of words'. Sinclair (1991: 113) calls this phenomenon 'a progressive delexicalization', or 'reduction of the distinctive contribution made by that word to the meaning'. Take, the phrase *take/have a walk*, for example. The frequent verb *have/take* has lost much of its meaning and the meaning is spread across the whole phrase rather than being restricted to one word or another. Sinclair (1994) discusses the word combination of *naked eye* to illustrate this phenomenon. Based on the concordance to *the naked eye* from the 1997 Bank of English corpus, Sinclair observes the following patterns of this combination: (1) *naked eye* typically appears at the end of a clause and is preceded by *the* (e.g., *Easily visible to the naked eye,...*), (2) *The naked eye* is preceded by *to*, or less frequently, *with* (e.g., *as a transformation invisible to the naked eye,...the base you can't see it with the naked eye but you know it's there...*), (3) The prepositional phrase *to/with the naked eye* follows a range of words related to sight, most frequently *visible* or a form of the verb *see* (e.g., *its effect cannot be seen by the naked eye...*), (4) About half the instances of this typical usage are negative (as against a general figure of 10% of all clauses being negative, see Halliday 1993b). Sinclair's conclusion is that there is a unit of which *naked eye* is a part, which has a specific meaning but a range of lexical realization. This unit does not correspond to any syntactic unit and Sinclair calls it a 'meaning unit'. If language is to be analyzed according to the idiom principle, the meaning unit would be the primary unit of analysis.

Sinclair's approach to collocation is not only revolutionary methodologically but also has helped us to obtain new insights on the nature of language. The idiom principle, as one of the central organizing feature of language, has been neglected for so long a time. The theoretical implications for constructing grammar out of lexis have only just begun to be understood. The concept of 'meaning unit', although far from being without problems, has pointed out a new direction for linguistic research, for which meaning has always been the main pursuit.

2.6 Summary

The study of collocation as a lexical level by Firth half a century ago has opened up a fruitful area of linguistic research. Firth's collocation study should always be seen against the background of his theory of meaning. It is a reaction against the longstanding referential approach to linguistic meaning. It is one of his considerations of the syntagmatic aspects of a lexical item's meaning potential and also one aspect of his interest in the predictability in discourse. It also reflects his general philosophical that language is by nature going-on. After prosodic analysis, collocation is the topic that provided the most interest for other scholars who had worked with him or been indirectly influenced by him. Among them are Mitchell, Halliday and Sinclair. Although they have different theoretical perspectives, the three scholars we reviewed in this section share one common point: they all consider collocation and other prefabricated language forms simultaneously. This tradition lays the foundation for us to investigate the functions of collocation and idiomatic expression at the same time.

Linguistic Prefabrication

A Discourse Analysis Approach

Ding, J.

2018, XIII, 158 p. 13 illus., Hardcover

ISBN: 978-981-10-7009-9