Definition of collocation : descriptive studies vs.semantic studies

KOYA, Taeko

Definition of collocation
— descriptive studies vs. semantic studies —

Taeko Koya

1. Introduction

Collocation has become one particular area of vocabulary research which has attracted attention theoretically and practically as vocabulary has been treated as a mainstream topic from the standpoint of language use rather than language structure since the 1980s. Many researchers (Benson et al., 1997; Bahns & Eldaw, 1993; Bahns, 1993; Channell, 1981; Lewis, 1993; Takizawa, 1999; Willis, 1990) have discussed the importance of collocation in terms of productivity, and they agree that learners must master how words combine or collocate with each other in order to develop their fluent and appropriate language use.

While a lot of collocation research has been carried out to date based on the agreement on the view of collocation mentioned above, theoretical studies on collocation have been insufficient in defining the concept of collocation in a rigorous way. The well-known statement "the tendency of a lexical item to co-occur with one or more other words" has been followed since the definition of collocations was first introduced by Firth (1957) to more recent and up-to-date research projects by many researchers (Backlund, 1976; Cruse, 1986; Crystal, 1992; Halliday, 1966; Ridout and Waldo-Clarke, 1970; Seaton, 1982). However, it does not explain the boundary between three phraseological concepts: free combination, idiom and collocation which can be presented along a continuum. Some semanticists have attempted to point out some shortcomings of the collocation studies by Firth and his followers, but they have not given strong alternatives to it.

The aim of this paper is to clarify how collocation studies have been

developed from the Firth's definition in the descriptive studies and how semanticists have tried to provide the new definition of collocation in terms of their framework. Although researchers in both domains started collocation with the assumption that words receive their meaning from words they co-occur with, they have described it from their perspectives. I believe that it is necessary to review collocation studies in two different research groups to formulate a comprehensive concept of collocation.

2. Previous research on collocations in the descriptive studies

2.1. Collocation studies in the Firthian concept

Collocation studies can be traced back to the work of H. E. Palmer, who is said to be the first linguist to use the term *collocation* in the present-day sense. In Palmer's *Second Interim Report on English Collocations* (1933) and *This Language-Learning Business* (1969), he defines collocation as word combinations which contain one or more words having meanings only in that collocation. This completely hinders learners from acquiring the whole range of collocations from weak collocations (e.g. *see a film* and *an enjoyable holiday*) to the strongest and most restricted collocations (e.g. *see reason* and *burning ambition*)

In both of his articles, he highlighted the importance of collocation in the successful language learning. He stressed this because some English teachers fail to realize the existence of abundant collocations, especially irregular collocations, which consist of semantically opaque constituents of words such as *foot the bill*. In order to acquire a wide range of collocations, he points out the necessity for both teachers and students to be aware of the need for collocation acquisition.

Although Palmer pioneered collocational research, Firth's definition of collocations (1957) and his stance that lexical studies in terms of syntagmatic aspects are important and that collocations play an important role in word's meaning had a greater influence on his followers' studies on collocations. His main concern is literary stylistics, where it is necessary to recognize the

distribution of words and certain collocations in order to explain literary effect. His definition of collocations is deduced from many examples of literary works sharing common classical sources. He explains collocations as follows (1957):

> The statement of meaning by collocation and various collocabilities does not involve the definition of word-meaning by means of further sentences in shifted terms. 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*. This kind of mutuality may be paralleled in most languages. (p. 196)

In other words, collocation refers to co-occurring associations of two or more lexemes in a specific range of grammatical constructions and "mutual expectancy of words" (p. 196) is a distinguishing feature of collocations.

Firth also describes two ways of investigating collocations (1957, p. 196): (a) the stylistics of what persists in and through changes, and (b) the stylistics of personal idiosyncrasies; and both of these influence the followers' studies, like Sinclair (1966, 1991).

According to Bollinger and Sears (1968), who follow the Firthian concept of collocations, the syntactic and habitual association of words, "the characteristic company they keep" (1968, p. 53) with external reality is regarded as collocation. They explain the process of word acquisition by using *holophrastically* and *horizontally*. First, a child apprehends the verbal expression *holophrastically*: In other words, he/she expresses an undivided word representing a total context. After that, he/she can differentiate larger combinations of words horizontally, in more extended pieces of speech based on the syntax. They point out that collocations are acquired relatively later through the process of L1 vocabulary acquisition.

Replying on Firth (1957), who stresses the importance of lexical studies in descriptive linguistics and regards the statement of collocations as the most

useful approach, Halliday (1966) is concerned with how collocation defines membership of lexical sets, which are "groupings of members with like privilege of occurrence in collocation" (1966, p. 153) and proposes the lexico-grammatical system in his own framework as follows:

Table 1. Lexico-grammatical system by Halliday (1966, pp. 152-153)

	grammar	lexis
paradigmatic axis	system	set
syntagmatic axis	structure	collocation

According to his framework of a language system, collocation is one of the main components and central part in his lexico-grammatical system. He defines collocation as a syntagmatic relation of words which is linear co-occurrence together with some measure of significant proximity (1966, p. 152) and says that collocational relations intersect with structural ones. He points out that grammar does not always give an explanation of the relations between *strong* and *powerful*. In fact, *strong* and *powerful* are both members of a set of items and both of them collocate with *argument*, but they cannot always collocate with the same words. While *strong car* and *powerful tea* will either be rejected as ungrammatical (or unlexical), *strong tea* and *powerful car* can be acceptable. This explains that both of them depend on the syntagmatic relation into which each enters but that collocational patterning is independent of grammatical structures (1966, p. 150).

In relation to the notion of collocation and lexical set, Malmkjaer (1991) explains that it can be exploited in decision making on polysemy and homonymy. The different collocational environments associated with the word *bank*, (e.g. *cheque, deposit, manager, money,* etc.) and *bank* (e.g. *river, water, earth, trees,* etc.) show that bank is a homonym (Malmkjaer, 1991). In the meantime, non-cognate word forms (e.g. *city* and *urban*) have the same collocations. Therefore, it can be shown that they belong to the same set (1991,

p. 303).

Greenbaum (1970) and Mitchell (1971) also follow the Firthian concept of collocations. They study collocations from an integrated approach of lexis and grammar and define them as association of words whose lexical and syntactic patterning is viewed as distinct but interrelated. They emphasize that in collocation study, the lexical view should not be separated from the syntactic view, because the collocational meaning is changed according to different syntactic patterning (Greenbaum, 1970, p. 12; Mitchell, 1971, p. 50).

Greenbaum (1970, p. 12) claims that limited, homogeneous grammatical classes yield the most useful analytic results and points out that a serious disadvantage of an item-oriented approach in the collocation study is that ignoring syntactic restrictions on collocations leads to inaccuracy. In other words, both lexical and syntactic patternings are important to fix collocations. If either is incorrect, then the phrase product is not a collocation. Greenbaum exemplifies the advantage of interrelated patterning between lexis and syntax by using an item *much*:

Much collocates with *like* in negative sentences, not in affirmative sentences.

I *don't like* him *much*.

But not   *I *like* him *much*.

However, if pre-modifiers are added to *much* in affirmative sentences, it can be acceptable:

I *like* him *very much/ too much/ so much*.

(Greenbaum, 1970, p. 12)

Mitchell (1971) has the same opinion as Greenbaum and maintains that the meaning of collocations is influenced not only by their contextual extension

of lexis but also the generalized grammatical patterns within which they appear. He focuses on morphemes of words and explains that words which contain the same lexical morphemes do not necessarily mean the same when rearranged or inflected. For example, *hard* in *hard work* means something different from *hard* in *hard-working.*

In addition to explaining collocations and illustrating them with some examples, McIntosh (1961, p. 193) proposes a useful framework of four categories for the determination of style in language as follows:

There is the possibility of four obviously distinct stylistic modes: normal collocations and normal grammar, unusual collocations and normal grammar, normal collocations and unusual grammar, unusual collocations and unusual grammar. (p. 193)

He argues that "normal collocations are too familiar and thus banal and abnormal collocations are unfamiliar and thus indecipherable" (1961, p. 193). In order to bridge the gap, standard language norms are necessary, because without this, it becomes impossible to communicate, as no one has the same experience or set of associations.

McIntosh also claims that native speakers of English are the only people who can produce new word combinations, using their intuition of the collocational range of words. He describes their intuition of new formations as "range-extending tendencies" (1961, p. 193). An example of collocation which has been created in this way is the lexical item *key* which has recently considerably extended its range: *key move, component, policy,* etc. Collocational acceptability including new formations can be statistically analyzed on large corpora which are the product of native-language speakers.

However, this argument of McIntosh is invalid in the background that English is now regarded as an international language around the world. Crystal (2003) updates Kachru's (1985) model showing the English-speaking population and explains as follows: In the Inner Circle, 320 to 380 million people

use English as a mother tongue; in the Outer Circle, 300-500 million people use English as a second language; and in the Expanding Circle, 500 million to one billion people use English as a foreign language. This figure indicates many non-native speakers around the world communicate with each other in English. Therefore, there is a great possibility that new word combinations spoken by not only native speakers but also non-native speakers of English can be mutually intelligible and acceptable.

Carter (1987, pp. 36-57), who is like a bridge between Greenbaum (1970), Mitchell (1971) and McIntosh (1961), fundamentally follows Firth's definition of collocations. He describes them as a group of words which co-occur repeatedly, and studies these patterns of co-occurrence grammatically and lexically. Grammatical studies look at the needs of particular pedagogic projects for ESL or EFL learners, while lexical studies aim to find the lexical patterning of texts and tend to make use of computerized lexical research. Carter emphasizes that because both grammatical patterning and lexical patterning influence the meaning of collocations, the study should include both of them, as do Greenbaum (1970) and Mitchell (1971).

Carter also explains collocation in terms of *frequency* and *range* as does McIntosh (1961) and above all, he stresses that the latter is the most suitable tool to describe the collocational restriction. He shows the difference of ranges between words, using some examples which shed light on his idea. Good examples he gives are *putrid, rotten, rancid* and *addled*, which have restrictive ranges and refer to a substance which has decayed and can no longer be used. *Putrid* collocates with fish, *rancid* with butter, oil, lard, etc., and *addled* is confined only to eggs, while *rotten* can collocate with fish and eggs and also with fruit. Thus, these properties of *rotten* mean that by this criterion it is a core word and, by contrast, *putrid, rancid, addled* are less core. In short, synonymic relations between words can be usefully distinguished with reference to the different collocational ranges of the synonyms involved. He concludes that this aspect of collocation is a valuable and revealing one and is one of particular relevance to vocabulary in language teaching.

The focus of Jackson (1988) is different from those of Greenbaum (1970) and Mitchell (1971), who are interested in the explanation of collocations in terms of not only a lexical view but also a grammatical view. Jackson also differs from McIntosh (1961), who focuses on productivity of new collocations by native speakers of English. He highlights the difference between collocations and idioms and states that "collocation is not a fixed expression, but there is a greater than chance likelihood that the words will co-occur" (1988, p. 96). For example, in the sentence, "He had a false _____," the nouns that fill the slot in this structure might include *eye, nose, beard, expectation* or *passport*. Without the article in this structure, teeth and eyebrows could be added. If the subject of the sentence were *the car* rather than *he, numberplate* might be expected. Thus, unlike idioms, collocations are combinations which regularly keep company not with one word but with certain other words.

Jackson also refers to a certain mutual expectancy, range and collocational restriction, which are important features in collocations. Especially, regarding mutual expectancy, which is reminiscent of Firth's definition, he exemplifies this using *tooth* and *false*. *Tooth* is more likely to collocate with *false* than *false* is to occur in combination with *tooth*, because a number of alternative nouns can be combined with *false* such as *eye, nose, beard, expectation* and *paper*, while tooth can be combined with fewer adjectives such as *irregular* and *decayed*.

Finally he points out (a) the importance of the corpus of spoken and written text and (b) the importance of lexicographers' intuitions and insights into their own and fellow-speakers' knowledge of language use in compiling dictionaries, and (c) criticism for the different state of treatment of collocations in some dictionaries (1988, pp. 99-103).

Aisenstadt (1979) is also concerned with describing collocations, compared with idioms and free combinations and considering certain characteristics of them. He introduces four characteristics. One is that restricted collocations have various structures (e.g. *V-N, A-N* and *V-Adv*) and the patterns are command admiration, attention, decision and so on. The second is that the meaning of restricted collocation constituents may be characterized by one of

the following: (a) the constituents have a very narrow specific meaning which does not allow a wide range of commutability (e.g. *shrug one's shoulders*), (b) the constituents are used in a secondary meaning, often abstract or figurative (e.g. *clench one's hands, clench one's fists*), or (c) the constituents have a weakened and grammaticalized meaning of verbs, which can be changed result in a possible interchange of those verbs (e.g. *give a laugh, have a laugh*). The third is that the commutability between the restricted collocation constituents may be restricted in the following: (a) both constituents of restricted collocations are restricted to a limited number of co-occurring words (e.g. *shrug one's shoulders, square one's shoulders, hunch one's shoulders*), or (b) the commutability of only one constituent of the restricted collocation is restricted (e.g. *make a decision, take a decision*). The fourth is that while many restricted collocations belong to the *neutral* layer of vocabulary, a certain number is mainly *colloquial*.

The interest of Halliday and Hasan (1976) is related to lexically predictable collocational chains that extend beyond the boundaries of a sentence in textual cohesion, different from those of other researchers who pay attention to idiosyncratic and unpredictable co-occurrences of words. They define collocation or collocational cohesion as all the various lexical relations which do not depend on any systematic semantic relation but which have the tendency to share the same lexical environment such as *bee...honey* and *walk...drive*. This is simply a cover term for the cohesion in which lexical items are associated with each other in some way or other.

2.2. Collocation studies from the synthetic angles

Collocations are systematically defined from multi-criteria by Gramley and Pätzold (1992) and Nation (2001). It seems that each concept of collocations is slightly different from that of earlier researchers in terms of the synthetic description. Gramley and Pätzold (1992) and Nation (2001) maintain that only one criterion is not enough to define collocations and that "the large number of scales needed is evidence of the range of items covered by the term" (Nation, 2001, p. 329).

Gramley and Pätzold (1992) define collocations as "combinations of two lexical items which make an isolable semantic contribution, belong to different word classes and show a restricted range" (1992, p. 61). This definition is based on six main criteria, which are discussed in detail below:

1. Two lexical items, not grammatical ones

This refers to what is defined as collocations mentioned above.

e.g.) strong coffee, white currant

2. Two categorizations: lexical combination and grammatical combination

In accordance with Benson et al. (1986), collocations can be divided into lexical combinations and grammatical combinations.

- Lexical combination — dominant words only (noun, adjective, adverb and verb)

  e.g.) compose music, strong tea, affect deeply

- Grammatical combination — dominant words + preposition or a grammatical construction

  e.g.) by accident, apathy towards, angry at

3. Independent meaning of constituents

This means that the "individual constituents contribute to the meaning of the combination as a whole" (p. 62). In other words, each constituent of a collocation can have special meanings which are restricted to one particular collocation, and the more opaque some individual constituents are, the more closely they are linked to the other constituents. This important point about collocational meaning helps to set up two different classes: collocations and idioms.

e.g.) white paint, white grape < white lie (= harmless), white night (= sleepless) < white horses (= white-topped waves), white coal (= water as energy source) ('<' means more opaque)

4. Word classes

Lexemes belong to different word classes such as *demand-meet* (noun-verb), *face to face* (noun-preposition-noun) and *apologize-profusely* (verb-adverb).

5. Range

This means that items are different in terms of their close relationship to other items and this criterion helps to distinguish between free combinations and collocations. In the case of *decide on a boat*, if the meaning is *choose (to buy) a boat*, then it contains the collocation *decided on*, while if the meaning is *make a decision while on a boat*, it is a free combination (cited in Benson et al., 1997). The number of free combinations is limitless and *the BBI dictionary* (Benson et al., 1997) is made up of only collocations.

6. Fixedness

Different collocations have different degrees of fixedness.

(a) Morphology

In some collocations, adverbs do not form the -ly morpheme such as *swear-blind* and *forget-clean*.

(b) Substitutability

In some collocations, constituents can be replaced by their synonyms. For example, *hardened criminal* and *confirmed criminal* are acceptable, although **hardened burglar* or **hardened murderer* are not.

(c) Additions and Deletions

Additions have taken place frequently, while deletions are much rarer. In case of additions, most often pre- or post-modifying nouns are normal.

e.g.) The oil-exporting nations...may soon restrict production below the level needed to *meet* still rising world *demand*. (demand-meet; B. Ward, *Progress for a Small Planet*, p. 15)

(d) Displacement

Personal pronouns can be substituted for constituents in some collocations:

e.g.) Quality is our *promise*. Cancellation is your privilege if we fail to *meet*
 it. (advertising material, *The Economist*, May 1991)

(e) Separability

In some collocations, word combinations, which belong to bound
collocations (Cruse, 1986, p. 41) can not be separated, such as *foot the*
bill and *curry favor*.

(f) Distribution

The distribution in collocations is relatively changeable. For instance,
They met their demands; their demands, which were not met completely...
are both acceptable.

As mentioned above, many criteria have to be met in order to regard word
combinations as collocations. Nevertheless, it can be said that collocations
themselves are vague because of a low degree of formal fixedness in
combinations with the composite semantic structure. Gramley and Pätzold
(1992) argue that one solution to distinguish between collocations and other
word combinations is to rely on one of objective criteria, the frequency of co-
occurrence of words in corpus-based research.

Nation (2001) argues that using a set of scales is the most effective way of
setting up criteria for grouping items as collocations, and setting up the groups
of collocation and 10 scales which have been identified by many researchers
would be needed to do it. Compared with the criteria Gramley and Pätzold
(1992) present, the criteria of Nation (2001) are further segmentalized,
including criteria 3, 4, 5 and 6 of Gramley and Pätzold (1992). Nation's 10
criteria are as follows:

1. Frequency of co-occurrence

A very important criterion is this frequency of co-occurrence, which
should be considered along with collocation range and which is measured by
computers in large corpora.

2. Adjacency

Collocates can occur from next to each other or be separated by variable words or phrases like an example, *"little did x realize"* (Renouf & Sinclair, 1991).

3. Grammatically connected

Collocates can usually be seen within the same sentence as part of a grammatical construction, but it is possible to see items within the same text, not grammatically connected to each other but in a lexically cohesive relationship as collocates. Kennedy gives such an example[1] as: "Her uniform was of rich raw *silk* in a *shade* which matched her hair" (1998, p. 113). In this sentence, *silk* and *shade* can be regarded as collocates in a lexically cohesive relationship.

4. Grammatically structured

In addition to *habitual co-occurring* of words, another criterion, *grammatically structured* should be needed. For example, *although he* and *of the* should not be contained in collocations which take account of the major divisions that would be made in analyzing a clause, although they occur so often (Kjellmer, 1982).

5. Grammatical uniqueness

Collocations range from grammatical uniqueness (e.g. *hell for leather*) to grammatically regular patterns (e.g. *weak tea*) with patterned exceptions like *go to bed/town/hospital* (without an article) as the mid-point.

6. Grammatical Fossilization

Collocations range from no grammatical variation (e.g. *by and large*) to changes in part of speech (e.g. *Her heart wasn't very strong and her life assurance premiums weren't cheap. It cannot have been easy to meet them.* [cited in Gramley and Pätzold, 1992, p. 65]) with inflectional change (e.g. *He kicked the bucket.*) as the mid-point.

7. Collocational specialization

Collocations range from always mutually co-occurring (e.g. *commit suicide*) to all occurring in a range of collocations (e.g. *hocus pocus*) with one bound item (e.g. *kith and kin*) as a mid-point. Collocational specialization is equivalent to what Aisenstadt (1981) calls *restricted commutability*.

8. Lexical fossilization

Collocations range from unchangeability (e.g. *No fear!*) to allowing substitution in all parts (e.g. *last month*) with allowing substitution in one part (e.g. *permit/allow/give access to* [cited from Schmitt, 2000, p. 79]) as a mid-point. These criteria indicate that collocations from this entire range should be included when counting the frequency of collocations.

9. Semantic opaqueness

Collocations range from semantic opaqueness (e.g. *of course*) to semantic transparency (e.g. *open the door*). This criterion and grammatical fossilization are commonly used to define an idiom.

10. Uniqueness of meaning

Collocations range from only one meaning (e.g. *on behalf of*) to several meanings (e.g. *kick the bucket* has two meanings: to die and to kick the bucket with your foot) with related meanings as the mid-point.

Nation points out that the ranges in each of the 10 criteria have all been graded from most lexicalized to least lexicalized and gives the example of *hocus pocus* as a highly lexicalized collocation according to the 10 criteria. However, criteria 1 and 2 are not regarded as relative scale, but absolute scale by computational researchers who try to indicate objective criteria such as counting and calculating frequency and range with statistics and computer techniques to identify collocations.

3. Previous research on collocations in the semantic studies

3.1. Collocation studies in the semantic framework and syntagmatic lexical relations

Semanticists criticize the studies of collocations by Firth and his followers as insufficient and attempt to investigate collocations in terms of semantic framework and syntagmatic lexical relations under the scope of semantics. In spite of mentioning some shortcomings of the collocation studies by Firth and his followers, they do not give strong enough support for their concept and the function of collocations.

Lyons (1966), Palmer (1976), Katz and Foder (1963) and Lehrer (1974) criticize Firthian studies of collocations and attempt to explain them in the semantic field.

Lyons (1966, pp. 289-297) considers Firth's treatment of collocations as insufficient in that he only places collocation in an intermediate level between grammar and situation, but he never defines the notion of collocation in his general theory. Firth (1957, p. 196) states that "one of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night*." Moreover, Firth introduces and exemplifies the notion of collocation and gives an answer for the question, "what collocations are synchronically acceptable or unacceptable?" by an analysis of Swinburne's poetic diction and of certain letters of the 18th and early 19th centuries. One of the findings in his analysis is that certain collocations remain current over long periods of time, whereas others do not. However, Lyons thinks that the analysis is not synchronic nor should the collocations shown by Firth be considered as collocations of units but as idioms. He suggests that collocations should be studied as part of the synchronic and diachronic analyses of language. Unfortunately, Lyons provides only criticism of the Firthian theory of meaning without any concrete alternatives.

Similar to Lyons (1966), Palmer (1976) also criticizes Firth's treatment of collocation as merely part of the meaning of a word and his limited concern with the more obvious and more interesting co-occurrences, the mutual expectancy of words. His argument as a semanticist is that there is a restriction on the use

with a group of words that are semantically related, which is the matter of range suggested by McIntosh (1961). He also states that there are three kinds of collocational restrictions: collocations based wholly on the meaning of the item as in the unlikely *green cow*, collocations based on the range which is characterized as some semantic features in common as in the unlikely *the rhododendron passed away* and collocations based on collocational restrictions in a strictest sense, with neither meaning nor range as *addled* with *eggs* and *brains*.

Katz and Fodor (1963, pp. 172-173) argue that a semantic theory of a language would "take over the explanation of the speaker's ability to produce and understand new sentences at the point where grammar leaves off" (pp. 172-173), and try to establish the semantic theory which would organize and systematize facts about meaning of lexical items which are perceived as a combination of the semantic properties, free from grammar. They also attempt to explain collocations with these semantic concepts of theirs. They are reflected in a language dictionary, in which words contain a selection restriction and can be explained with a semantic marker. In this way, they seem to consider that the semantic approach is more likely to explain why words can be found together.

Lehrer (1974, pp. 173-176) provides a more comprehensive explanation for insufficiency of the collocation studies by Firthian linguistics. From his semantic viewpoints, he points out some contradictory points in the lexical hypotheses whose stance is taken by Firth (1957) and Halliday (1966) as follows:

1. Firth never gives an exact definition or paraphrase of collocational meanings. In his 1957's paper, he mentioned that "meaning by collocation is an abstraction at the syntagmatic level" (cited in Lehrer, 1974, p. 174) and is not directly concerned with the conceptual approach to the meaning of words. Although he gives an example of *night* and *dark* as a collocation, he never gives a proper definition.

2. Halliday has attempted to discuss the notion of a collocational level in terms of a general theory of grammar, but he failed. There are two other alternatives. One is that formal linguistic patterns can be explained with not only grammar but also semantic terms. The other is to recognize that there will still remain patterns which cannot be accounted for in formal linguistic patterns even after a grammar has been constructed.

3. Halliday recognizes that frequency of occurrence is necessary in collocation study, but he does not focus much on study of large corpora as a source of data on collocation and sets. It is important to notice which pairs of words occur with less than expected frequency, as well as those that occur with more than expected frequency.

4. Sinclair (1966) discusses a lot of theoretical and methodological problems involved in the study of text to discover collocational sets, but he includes deviant sentences such as poetry and fiction stories. It is important to use all kinds of discourse, not specific one in determining collocational sets.

5. Frequency studies are exaggerated as a useful way of determining what words belong together in a lexical set, but the more important question is what conclusions are to be drawn from the results. Furthermore, there is no explanation of why certain pairs of terms occur less often than expected.

Thus, Lehrer (1974, pp. 173-176) criticizes the lexical hypotheses in which Firth and other researchers cannot explain selection restriction in regard to the above five points.

Then he attempts to explain collocations with semantic features, but it seems to be extremely ambitious. Considering the direct object of *smell*, [smellable] is useless unless every possible direct object for *smell* in the language can be covered with the feature [smellable]. For example, possible direct objects are more than physical objects like *fresh air* or *stale smoke*. If they

are forced to be classified into physical objects, we have to accept the following examples, "*He hit the fresh air with a stick*" and "*I smelled a rat*", but we cannot.

He also referred to arguments of Leisi (1953), Gruber (1965) and McCawley (1970) who attempt to explain semantic restrictions with semantic features. They all exemplify paradigmatic contrasts involved in words so that differences of selection restrictions can be predicted on the basis of meaning and attempt to devise some features that would describe the rest of the cases. For instance, *kick, slap* and *punch* McCawley (1970, pp. 180-181) raised as examples refer to specific actions as well as to the instruments *foot, open hand* and *fist*. *Kick* can take as an instrument what can be put on or in some way attached to a foot like a *He kicked me with a boot (slipper, skate, snowshoe, ski)*. *Slap* needs an open hand, a folded newspaper or even a fish (e.g. *I held the fish between my toes and slapped the cat with it when she tried to eat it.*), while *punch* means to hit a compact, hard and physical object with one's fist (e.g. *He punched the wall angrily, then spun round to face her.*) Although Lehrer (1974) admits McCawley's point of view that different words have certain distributable meanings, whether it is correct or least correct for most people remains to be worked out.

In conclusion, he proposes the mixed positions in which the lexical and semantic positions on lexical co-occurrence are combined, but he does not support any strong evidence for why the mixed position is possible.

Cruse (1986) considers collocations as one of the most important components in the semantic field. Cruse (1986) defines collocation in terms of three important points: frequency, collocational restriction and semantic opacity. According to Cruse (1986, p. 41), collocation is regarded as sequences of lexical items which habitually co-occur, which are highly restricted contextually, but which are basically transparent in the sense that each lexical constituent is semantic. Compared with idioms, collocations are semantically more transparent: however they have some distinctly idiom-like characteristics, too. For instance, *foot the bill* and *curry favor*, are not likely to be separated.

Cruse, who is similar to other semanticists, also describes collocational

restrictions as "semantic and arbitrary co-occurrence restrictions" (1986, p. 279), by examining the syntagmatic meaning relations between lexical units. For example, *kick the bucket* is only used with a human subject, but its propositional meaning is merely *die* and not *die in a characteristically human way*. That is, "the restriction to human subjects is semantically arbitrary" (1986, p. 279). Then, the collocational restrictions are divided into three kinds from the viewpoint of the degree to which they can be set out in terms of necessary semantic traits. One is systematic collocational restrictions when they can be fully specified. For example, *grill* and *toast* denote the same process or action from the view of *agents* but different from the point of view of *patients*. They are different in the method that we *grill* raw food and we *toast* cooked food. The second is semi-systematic collocational restrictions when some collocations have certain exceptions. For example, *customer* gets something material in exchange for money, whereas *client* typically receives less tangible professional or technical service. The last one is idiosyncratic collocational restrictions when their collocational ranges can only be illustrated by listing allowable collocates. For example, it is possible to say *spotless kitchen* but not to say *flawless kitchen*.

3.2. Collocation studies in the semantic prosody

Since the 1990s, some researchers have been interested in semantic prosody, which is introduced by Sinclair (1991) and Louw (1993). It refers to "a consistent aura of meaning with which a form is imbued by its collocates" (Louw, 1993, p. 157) or "a standard distinction between aspects of meaning which are independent from speakers (semantics) and aspects which concern speaker attitude (paradigmatics)" (Stubbs, 2001, pp. 65-66)[2]. *Prosody* originated in phonological coloring which is able to go beyond segmental boundaries and Sinclair applies the term *prosody* to semantic features of collocations whose habitual collocates are capable of coloring them. As a corpus is being developed, semantic prosody, which is inaccessible through intuition, can easily be obtained through the objective examination of how language is actually used via computers. In other words, corpus linguistics reveals a greater mismatch

between the products of introspection about language and those of extrospection and new objective language facts about language. Thus, collocation has been studied within the new semantic domain with the assistance of corpus linguistics.

Sinclair (1991) strongly argue that how carefully language is patterned can be gained by selecting text and considering all the instances by showing concrete examples. One example is a phrasal verb *set in* and by the search of the COBUILD corpus 114 examples are examined. His main finding is that *set in* commonly collocates with unpleasant states of affairs and only three refer to the weather; a few are neutral, such as *reaction* and *trend*. What typically sets in is *bad weather, decay, despair, rot* and *rigor mortis* and not one of these is conventionally desirable or attractive. He recommends that building up these kinds of database for teachers' reference by accessing much more reliable information from corpora leads teachers to provide a more confident way of teaching to learners, although he shows no concrete materials for direct exploitation in the classroom.

Louw (1993) explores the possibility of semantic prosodies and suggests what role they may play in texts. In order to examine semantic prosodies of *utterly* which are found in Larkin's poems, 99 citations drawn from the original 18 million word corpus at COBUILD are analyzed. The concordance shows that *utterly* connotes *bad* as in *utterly confused* and *utterly ridiculous*. On the other hand, there are few *good* examples. In fact, only four examples are found as *good* but all of them carry a fairly obvious ironic intention. For the findings he makes a comment that they enable us to determine criteria for recognizing semantic prosodies, only because "the prosody on *utterly* is as consistent as it is that it admits the possibility for irony" (Louw, 1993, p. 164). Thus, large corpora allow us to extract profiles of semantic prosodies which reveal the real speaker's intention.

The aim of Stubbs' study is to demonstrate semantic prosody of lexical collocation, which is one of the norms of the use by investigating corpora, along with other researchers such as Sinclair (1991) and Louw (1993). He regards

collocation as "the habitual co-occurrence of words and a purely lexical relation between words in linear sequence, irrespective of any intervening syntactic boundaries" (1995, pp. 245-246). In his later study (2001, p. 29), he clearly adopts the rather statistical stance that collocation should be defined as frequent co-occurrence. He argues that in order to identify lexical collocations, semantic prosody should be examined by studying corpora of naturally occurring data. He gives a brief example of the semantic field of *cause and effect* by accessing corpora. Table 2 shows his explanation and examples.

Table 2. Semantic field of *cause and effect* by Stubbs (1995, pp. 252-253)

words	semantic prosody	collocates
CAUSE	predominantly negative	*accident, problem, disease, chaos, embarrassment*
CREATE	neutral	*condition, havoc, illusion, image, problems*
REASON	not very clear	*altruistic, apparent, cogent, compelling, different,earthly, good, main, obvious, political*
RESULT	neutral	*disappointing, end, expected, final, inconclusive, interim, preliminary, unintended*
AFFECT	very negative	*adversely, badly, directly, negatively, seriously*
EFFECT	very negative	*adverse, deleterious, devastating, dramatic, harmful, ill,negative, profound, toxic.*
CONSEQUENCE	very negative	*catastrophic, devastating, dire, disastrous, grave, negative*

Stubbs recommends that teachers use such semantic prosody that cannot be obtained without corpus assistance and that it is very reliable information about collocations for pedagogical implication.

Rudanko (2001) is also concerned with examination of semantic prosody. He claims that the concept of connotation should be investigated not only at the level of images aroused by individual lexical items, but at that of collocational patterns, and in order to examine it, not the intuition of native speakers but large electronic corpora is needed. He supports Sinclair's study of the concept of

collocational coloring in 1991, but points out the problem that his study is limited because only present-day English is dealt with. Then, based on these claims of his and the problem of Sinclair's study, he examines changes of the collocational coloring of the verb *set in*, of the adjective *bent* and *fraught* and of the verbs *cause* and *bring about* in three different corpora of collected examples from the 18th century, the 19th century, and the 20th century, for the three different centuries. The Chadwyck-Healey Corpus of 18th Century Fiction, the Corpus of 19th Century English, and the COBUILD corpus are respectively used for 18th century English, 19th century English and 20th century English. The findings show that there can naturally be different degrees of collocational coloring in each target word. For example, the two verbs expressing *causation*, *cause* and *bring about* are different in collocational coloring. In present-day English, *cause* is apt to take a negative complement, whereas *bring about* has a neutral or positive complement. In the 19th century, *cause* had a tendency to choose objects whose referents were ordinarily either neutral, or unpleasant in flavor, while *bring about* seemed less common in this century and tended to select objects which referred to events or things that were conventionally neutral or pleasant in flavor. In the 18th century, *cause* had objects that referred to events, actions or properties that had a negative flavor, while *bring about* was less frequent than *cause* in this century and it accompanied relatively various objects. Thus, from his diachronic study, collocational coloring is found to change according to different centuries.

4. Conclusion

This paper reviews collocation studies in the descriptive and semantic domains. In descriptive studies, it is Firth (1957) who first elaborated the theory of meaning in terms of syntagmatic aspects of lexis and explored the distribution of words in a text and how some occur predictably together with others. His notion of collocation has profoundly influenced his successors in Firthian studies, who further examine collocations based on his concept. However, over the years, collocations have been gradually defined from more

and more angles, using an increasing number of different features. This is necessary because collocations are impossible to describe in terms of only one feature. This reflects our increasing understanding of the complexity of collocations.

In semantic studies, in contrast to descriptive studies, semanticists, who have discussed lexical relations and types of meanings in terms of the paradigmatic aspect of lexis, criticize the studies of collocations by Firth and his followers as insufficient because of certain inadequacies. Then, they have attempted to investigate collocations in terms of the semantic framework and syntagmatic lexical relations under the scope of semantics. In spite of mentioning some shortcomings of the collocation studies by Firth and his followers, they do not give enough strong support for their concept and function of collocations in terms of semantic markers and so on. In the 1990s, in connection with the development of corpus linguistics, semantic prosody became a new concept to express "a consistent aura of meaning with which a form is imbued by its collocates" (Louw, 1993, p. 157) and some researchers have examined it.

Although researchers in both domains have developed collocation studies in their own framework, they have a similar feature in research approach: before the 1990s they mainly treated collocation in terms of the subjective criteria such as semantic opacity and collocational restriction and after the 1990s they relied more on the objective criteria such as frequency connected to be calculated with computers. In short, since the 1990s, studies in both domains have been influenced by computational studies connected with the development of computer technologies, which allow them to analyze and compare many materials such as individual texts and large samples of writing by computers. Thus, collocations should be defined from both of subjective and objective criteria in order to identify their framework more clearly.

Notes

This paper is based on the Chapter 2 from the author's PhD dissertation submitted to the Graduate School of Education, Waseda University, in 2005, under the title "The Acquisition of Basic Collocations by Japanese Learners of English."

[1] Kennedy's example seems to be wrong: silk collocates with not shade, but raw and rich. The example of distribution given by Gramley and Pätzold (1992) can be applied to grammatically connected.

[2] Stubbs (2001, pp. 65-66) preferably uses discourse prosodies instead of semantic prosody, in order to express the relation to speakers and hearers and their function in creating discourse coherence.

References

Aisenstadt, E. (1979). Collocability restrictions in dictionaries. In R.R.K. Hartmann (Ed.), *Dictionaries and their users: Papers from the 1978 B.A.A.L. Seminar on Lexicography* (pp. 71-74). Exeter: University of Exeter.

Aisenstadt, E. (1981). Restricted collocations in English lexicology and lexicography. *ITL: Review of Applied Linguistics, 53,* 53-61.

Backlund, U. (1976). Frozen adjective-noun collocations in English. *Cahiers de Lexicologie, 28* (1), 74-88.

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal, 47*(1), 56-63.

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System, 21*(1), 101-114.

Benson, M., Benson, E., & Ilson, R. (1986). *Lexicographic description of English.* Amsterdam: John Benjamins.

Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum, 1*(1), 1-14.

Bolinger, D., & Sears, D. A. (1968). *Aspects of language.* New York: Harcourt

Brace Jovanovich Inc.

Carter, R. (1987). *Vocabulary: Applied linguistic perspective*. London: Allen and Unwin Ltd.

Channell, J. (1981). Applying semantic theory to vocabulary teaching. *ELT Journal, 35*, 115-122.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.

Crystal, D. (1992). *A dictionary of linguistics and phonetics*. London: Andre Deutsch.

Crystal, D. (2003). *English as a global language*. Cambridge: Cambridge University Press.

Firth, J. R. (1957). *Papers in linguistics*. Oxford: Oxford University Press.

Gramley, S., & Pätzold, K. M. (1992). *A survey of modern English*. London: Routledge.

Greenbaum, S. (1970). *Verb intensifiers in English: A experimental approach*. The Hague: Mouton.

Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robinson (Eds.), *In memory of J. R. Firth* (pp. 148-162). London: Longman.

Halliday, M. A. K., & Hassan, R. (1976). *Cohesion in English*. London: Longman.

Jackson, H. (1988). *Words and their meaning*. London and New York: Longman.

Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: the English language in the outer circle. In R. Quirk, & H. G. Widdowson (Eds.), *English in the world*, (pp. 11-30). Cambridge: Cambridge University Press.

Katz, J. J., & Foder, J. A. (1963). The structure of a semantic theory. *Language, 39*(2), 180-210.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman.

Kjellmer, G. (1982). Some problems relating to the study of collocations in the Brown Corpus. In S. Johansson (Ed.), *Computer corpora in English language research*, (pp. 25-33). Bergen: Norwegian Computing Centre for the Humanities.

Lehrer, A. (1974). *Semantic fields and lexical structure*. Amsterdam: North-Holland Publishing Company.

Lewis, M. (Ed.). (1993). *The lexical approach*. Hove: Language Teaching Publications.

Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology* (pp. 157-176). Amsterdam: Benjamins.

Lyons, J. (1966). Firth's theory of meaning. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J.R. Firth* (pp.288-302). London: Longman.

Malmkjaer, K. (Ed.). (1991). *The linguistics encyclopedia*. London and New York: Routledge.

Matsuno, K., & Sugiura, M. (2002). *The definition of collocation*. Retrieved May 16, 2004, from http://sugiura3.gsid.nagoya-u.ac.jp/project/nnscollocation/report/ Matsuicollocation-sugi.pdf.

McIntosh, A. (1961). Patterns and Ranges. *Language, 37*, 325-337.

Mitchell, T. F. (1971). Linguistic "going on": collocations and other lexical matters arising on the syntagmatic record. *Archivum Linguisticum, 2*, 35-69.

Nakamoto, K. (1997). Another look at collocation: frequency, collocational restriction, and semantic opacity. *English Usage and Style, 14*, 58-67.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Palmer, F. R. (1976). *Semantics: A new outline*. Cambridge: Cambridge University Press.

Palmer, H. E. (1933). *Second interim report on English collocations*. Tokyo: Kaitakusha.

Palmer, H. E., & Redman, H. V. (1969). *This language-learning business*. London: Oxford University Press.

Renouf, A., & Sinclair, J. McH. (1991). Collocational frameworks in English. In K. Aijmer, & B. Altenberg (Eds.), *English corpus linguistics studies: Studies in honour of Jan Svartvik* (pp. 128-144). London: Longman.

Ridout, R., & Waldo-Clarke, D. (1970). *A reference book of English*. London: Macmillan.

Rudanko, J. (2001). *Case studies in linguistic pragmatics*. Lanham: University Press of America.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.

Seaton, B. (1982). *A handbook of English language teaching terms and practice*. London: Macmillan.

Sinclair, J. McH. (1966). Beginning the study of lexis. In C. E. Basell, J. C. Catford, M. A. K. Halliday, & R. H. Robins (Eds.), *In memory of J. R. Firth* (pp. 410-430). London: Longman.

Sinclair, J. McH. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Stubbs, M. (1995). Corpus evidence for norms of lexical collocations. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics* (pp. 245-256). Oxford: Oxford University Press.

Stubbs, M. (2001). *Words and phrases*. Oxford: Blackwell Publishers.

Takizawa, N. (1999). Collocation: Internet linguistic information 16. *Gengo 4* [Language], 96-97.

Willis, J. D. (1990). *The lexical syllabus*. London: Collins.