

# An Analysis of English Collocation

Emiko Takano

## Abstract

With the revolutionary progress in computer science, interest in vocabulary studies has been increasing for the last decade not only in the field of linguistics but also in applied linguistics. Making use of corpora has enabled researchers to analyze linguistic data objectively in a surprisingly shorter period of time. Collocation, which is defined as a group of words that occurs repeatedly in a language, is fundamental to the study of vocabulary. It is true that corpora have greatly helped research on collocation, however, researchers' judgment and analysis of patterning in collocation are truly needed.

This paper explores the concept of collocation and establishes a framework of collocation for analysis. There are still some subjects which have not been fully examined in this area, such as classification of collocation, clarification of multi-word units / items, and collocational network in our mental lexicon. There seems to be a great possibility for linguists to pursue collocation in the study of lexis with the headway of computer technology.

## Introduction

The study of vocabulary has been attracting more and more interest of researchers for the last decade or so (Tono 1997). This phenomenon, however, goes back to the late 1960s according to McCarthy (2001: 35): 'in the late 1960s... the paradigm of lexical studies began to shift in Western Europe, and more and more energy began to be devoted to collocational meaning in vocabulary.' Evidently, that 'energy' has been accelerated by the revolutionary progress in computers in the last 15 years.

Development of corpora, that is, electronic database of an enormous amount of linguistic data, has made it possible to observe data rapidly and objectively. This can be applied to compiling dictionaries. With the first edition (1987) of the Collins COBUILD English Dictionary as a start, almost all the dictionaries published in the U.K. have used a computer corpus in their compilation (Jackson 2002). Not only learners' dictionaries but native speaker dictionaries have also developed significantly owing to corpora.

It is true that large corpora together with effective software and powerful computers have greatly helped research on collocation. Nation (2001: 56), however, claims that 'studies of

collocation which have relied solely on computing procedures have yielded results which are not very useful.' Computers can give us information only to a certain point. It is researchers that make sense of the results processed by computer. What we need are researchers' judgment and analysis.

In this paper, we shall analyze collocation in the study of lexis. Section 1 explores the theoretical background of collocation, and Section 2 establishes a framework of collocation for analysis.

## 1. Theoretical background of collocation

Section 1 tries to explore the concept of collocation referring to several linguists' insights. It is necessary to clarify where collocation stands in the study of lexis. McCarthy (1990) points out two kinds of lexical relations: collocational relationships and sense relations such as synonymy, antonymy, and hyponymy. Sense relations are concerned with paradigmatic relations: the way that words are related to one another. On the other hand, collocational relationships are concerned with syntagmatic relations: the way that words combine with other words. Thus, collocation is defined as 'the habitual juxtaposition of a particular word with another word or words with a frequency greater than chance' in *The New Oxford Dictionary of English* (1998). Section 1.2. will further discuss the definition of collocation.

### 1.1. Tradition of collocation

This section mainly surveys the work of a group of British linguists in the 1960s and 1970s. The tradition of looking at lexis in the syntagmatic way was established by the linguists discussed below.

Let us first refer to the word *collocation*. It is said that the word was first brought into lexical studies as a technical term by Firth, although it was not originally Firth's. Nation (2001) points out that Palmer used the word *collocation* many years before Firth, claiming that collocation 'must or should be learnt, or is best or most conveniently learnt as an integral whole or independent entity, rather than by the process of piecing together their component parts' (Palmer 1933: 4). Firth (1957), on the other hand, referring to the syntagmatic tendency of words to work together, notes that part of a word's meaning is its collocations, that is, *dark* is part of the meaning of *night*, and vice versa. Thus, collocation is a central part of the meaning of a word in Firth's view.

Firth's concept of collocation was taken further by McIntosh (1966), who claims that a distinction should be made between grammar and lexis when looking at language, and proposes the term 'range' for the lexical patterning and 'pattern' for the grammatical patterning. Some words have a wider range, and others are restricted in range. McIntosh explains a range using an example *molten* as 'the fairly strictly limited inventory of nouns which may be qualified by

the word *molten*.' The range of the adjective *molten* would be *metal, rock, iron, lava*, etc. McIntosh makes further remarks that collocation ranges are not fixed as grammatical patterns. When we are confronted with an unusual collocation such as *molten feather*, we rely on our knowledge of the range and might dismiss it as unacceptable. Or we try to understand it by supposing an extension of range. This 'range-extending' tendency is considered as a characteristic of language. Carter and McCarthy (1988: 33) present an example, *software* in computer terms as 'an extension of the previously limited *hardware*.'

Halliday and Sinclair, who are considered as neo-Firthians, view lexis as an independent linguistic level. Halliday (1966) considers 'the nature of lexical patterns in language' and 'a lexical theory (which is) complementary to but not part of grammatical theory.' His account of lexis is said to be in parallel with grammar but as a separate and independent linguistic level as Table 1 shows.

Table 1. Lexis as a linguistic level

	syntagms: chain	paradigms: choice
grammar	structure	system
lexis	collocation	set

Sinclair (1966) agrees with Halliday (1966) on the basic points about collocation, regarding it as probability of co-occurrence. Sinclair focuses on collocation in a statistically significant way. McCarthy (2001: 62) points out that 'both Halliday and Sinclair in their early papers on collocation foresaw and spearheaded the development of the large-scale analysis of lexis using massive amounts of text.' Thus, the COBUILD project at Birmingham University led by Sinclair has completed a dictionary which is helpful to learn collocation. The COBUILD dictionary (1987) 'offers a lot of help in making this area (the patterns of collocation) clear' using the COBUILD corpus' (Carter and McCarthy 1988: 36).

As Halliday and Sinclair regard lexical study as separate from grammar, not all linguists agree with them on this point. Greenbaum (1970) and Mitchell (1971) take views that lexis and grammar are interrelated. It is worth noting that both Halliday and Sinclair have later modified their position on lexis as a linguistic level, showing a shift to an integrated approach. They have taken the position that 'there are crucial interdependencies between grammar, lexis and semantics.' (Carter 1998:62)

Greenbaum (1970) criticizes the item-oriented approach to collocation, 'which works independently of syntax and semantics' (Carter and McCarthy 1988: 37). Claiming that it obscures syntactic restriction on collocation, Greenbaum exemplifies that '*much* collocates with a preceding verb *like* in negative sentences' such as (a), 'but not in affirmative sentences' such as (b), however, sentences like (c) 'become perfectly acceptable if *much* is premodified.

- (a) I don't *like* him *much*.
- (b) \*I like him much.
- (c) I like him very / too / so much.

Mitchell (1971) mentions that 'lexical particularities are considered to derive their formal meaning not only from contextual extension of a lexical kind but also from the generalized grammatical patterns within which they appear' (48) and claims that making grammatical generalizations about collocation is important. As Mitchell suggests, it seems effective especially in language teaching to make lexical statements alongside grammatical generalizations.

### 1.2. Definition of collocation

*Collins COBUILD English Dictionary* (2001) defines the word *collocation* as 'the way that some words occur regularly whenever another word is used.' This definition seems insufficient because the groups of words such as *although he*, *but if*, and *of the* occur frequently in frequency counts of corpora (Nation 2001). It is true that a group of words, for example 'the car' frequently occurs in English. However, the word 'the' may collocate with any noun in the form of a definite article plus noun. Therefore, 'frequent co-occurrence' is not enough for defining collocation.

Nation (2001: 324) describes the second part definition of collocation as 'to contain some element of grammatical or lexical unpredictability or inflexibility.' Let us first take an example 'kick the bucket,' which means 'to die.' The meaning of the expression is unpredictable by adding the meanings of each word; *kick*, *the*, and *bucket*. Let us next consider the expression 'kith and kin,' which means 'family and friends.' This only occurs in the order of kith and kin, never in kin and kith. The order, or the form of the expression is inflexible. Thus, we shall introduce three keys to investigate collocation: frequency of co-occurrence, predictability, and flexibility.

### 1.3. Importance of collocation

Nation (2001: 318) introduces some arguments on collocation, which justify giving attention to word groups.

- (1) Language knowledge is collocational knowledge.
- (2) All fluent and appropriate language use requires collocational knowledge.
- (3) Many words are used in a limited set of collocations and knowing these is part of what is involved in knowing the words.

McCarthy (1990: 13) argues that collocational knowledge is useful to both native speakers

and learners of English.

Knowledge of collocational appropriacy is part of the native speaker's competence, and can be problematic for learners in cases where collocability is language-specific and does not seem solely determined by universal semantic constraints (such that 'green blood' would be odd in any human culture).

For non-native speakers of English, unpredictability of collocation is problematic. Semantic opaqueness and uniqueness of meaning give learners a heavy burden. Also the interference of L1 might cause a serious problem. In Japanese, for example, we use 'drink' instead of 'take' when expressing 'take medicine' as in '*kusuri wo nomu.*' However, as Firth (1957) noted that part of a word's meaning is its collocations, once we have learned the collocation 'take medicine,' we can apply it to other expressions such as 'take a pill,' 'take a tablet.' Therefore, collocation must be fundamental in the study of vocabulary because it is possible to gain part of the meaning of a word from the company it keeps.

#### 1.4. Classification of collocation

It is difficult to determine what should be classified as a collocation because collocations differ in size, type, closeness of collocates, and the possible range of collocates (Nation 2001).

It is often said that extremely fixed collocations are idioms, such as *kick the bucket* (Ikegami Ed. 1993). Kadota (Ed. 2003), presenting Schmitt's four level classification of collocation (2000) as in Table 2, claims that groups of words fallen into Level 1 are sometimes called multiword units / items (MWU / MWI). It is interesting to note that Schmitt's classification of collocation depends on the degree of frequency of co-occurrence and the degree of predictability / flexibility, which we discussed in Section 1.2..

Table 2. Classification of collocation

Level 1	idiom (frozen)	<i>kick the bucket / *kick the pail</i>
Level 2	fixed but transparent	<i>break a journey</i>
Level 3	substitution possible with limited choices	<i>give / allow / permit access to</i>
Level 4	two slots	<i>Get / have / receive a lesson / tuition / instruction</i>

Although this paper does not deal with clarification of the relationship between idioms and collocation, or MWU / MWI and collocation, it seems worth researching.

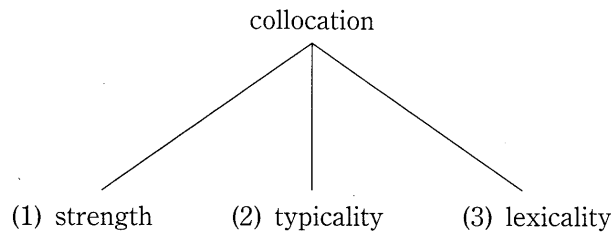
Nation (2001: 325), according to the two part definition of collocation: '(1) being closely structured and (2) containing some element of unpredictability,' presents 10 criteria to classify



Each of the ten scales has been graded from most lexicalized to least lexicalized. The word like *hocus pocus*, therefore, is a highly lexicalized collocation, which is represented on the very left side of each scale. Of course, there seem to be very few collocations which are high on every scale. Most collocations are high for only some of these scales. It is interesting to note that the three keys to investigate collocation: frequency of co-occurrence, predictability, and flexibility are somehow concerned with the above criteria.

## 2. Collocational framework

The previous section explored the concept of collocation. This section establishes a collocational framework for analysis. From the theoretical study of collocation in Section 1, I shall present the following framework for analyzing collocation.

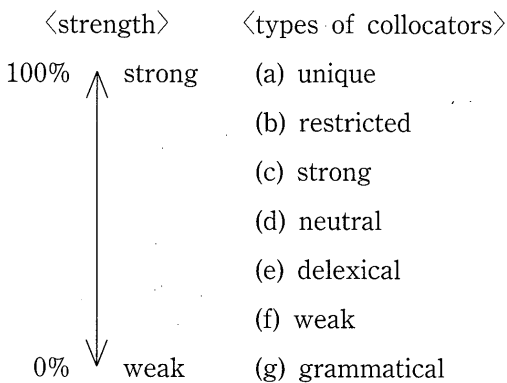


It is claimed that collocation can be analyzed in terms of strength, typicality and lexicality. The next three sections explain each constituent respectively.

### 2.1. Strength

It is said that any item can be measured by its collocability, the strength to combine with other words. This collocability can be represented on a scale.

#### A scale of collocability



Unique collocators have 100 percent collocability, in other words, their own environments

are 100 percent predictable. For example, *fro*, *spick*, and *kith* only collocate with *to*, *span*, and *kin* respectively. Therefore, 'to and fro,' 'spick and span,' and 'kith and kin' are regarded as single lexical items.

Items which have quite high collocability are called 'restricted collocators.' For instance, *blond* and *auburn* have extremely high probability to combine with *hair*. It can be claimed that they have a narrow range of collocation. Restricted collocators and strong collocators are sometimes treated within the same group.

As delexical collocators, general nouns and delexical verbs such as *have*, *take*, and *make* are pointed out. Halliday and Hasan (1976: 274) define general nouns as 'a small set of nouns having generalized reference within the major noun classes, those such as "human noun," "place noun," "fact noun" and the like.' Examples are:

[human] people, person, man, woman, child, boy, girl

[non-human animate] creature

[inanimate concrete count] thing, object

[inanimate concrete mass] stuff

[inanimate abstract] business, affair, matter

[action] move

[place] place

[fact] question, idea

As for delexical verbs, we have to understand their meaning by the words they keep company with since they have a broad range of collocations.

e.g. take a look, take time, take leave

make a noise, make a mistake, make an offer

have a think, have dinner, have an operation

Items which have low collocability are called weak collocators. As an extreme case, grammatical words such as articles, pronouns, prepositions are pointed out. They are least restricted collocationally.

## 2.2. Typicality

The notion of typicality is very important because knowledge of collocation depends largely on a question of typicality: what words are most likely to occur together. Some collocations are normal, but others are unusual. It should be noted that collocation can never be absolute. Native speakers would react to collocations as more acceptable or less acceptable on a scale of acceptability. Table 3 below shows three kinds of collocation according to typicality: normal (✓),



unusual but still acceptable (?), unacceptable (×).

Table 3. Collocational grid of 'adjective + noun'

	<i>problem</i>	<i>amount</i>	<i>shame</i>	<i>man</i>
<i>large</i>	?	√	×	√
<i>great</i>	√	√	√	√
<i>big</i>	√	√	×	√
<i>major</i>	√	?	×	×

(McCarthy 1990: 14)

When we judge collocations, what is crucial is 'typicality,' but not 'correctness.' Such typicality of collocation can be represented by 'markedness.' There are three types of markedness:

- (1) unmarked (e.g. a pretty girl)
- (2) marked (e.g. a smiling day)
- (3) deviant (e.g. sneezing bottle-opener)

The unmarked collocation is a 'normal collocation'. When we encounter 'a pretty girl,' we can immediately understand it without any reference. When we encounter 'a smiling day,' however, we may modify our view about 'smiling' and interpret it as 'a cheerful day.' In literary texts, we can find many marked collocations because writers often manipulate collocations to make their texts more imaginative and creative. Not only literary texts, but also tabloid newspapers or advertisements are full of marked collocations in order to catch readers' / listeners' attention.

Deviant collocations can be regarded as extremely unusual ones. This type of collocation may be found in literature, especially in modern poetry. However, the most likely text in which we can find a deviant collocation is said to be foreign learners' writing. For the learners, it is often a problem to produce typical collocations as well as recognizing untypical collocations. Even if we recognise it as a marked collocation, it is a problematic task to understand it including subtleties.

### 2.3. Lexicality

According to lexicality, collocation can be divided into three types: lexical (e.g. auburn hair), grammatical (e.g. the city of Tokyo), and lexico-grammatical (e.g. restricted only to 'adjective + noun'). Carter (1998) explains these three types using the example of the word *consent*.

lexical: the word *consent* co-occurs with 'mutual,' 'common'

grammatical: the verb *consent* is followed by the preposition *to* and another verbal or

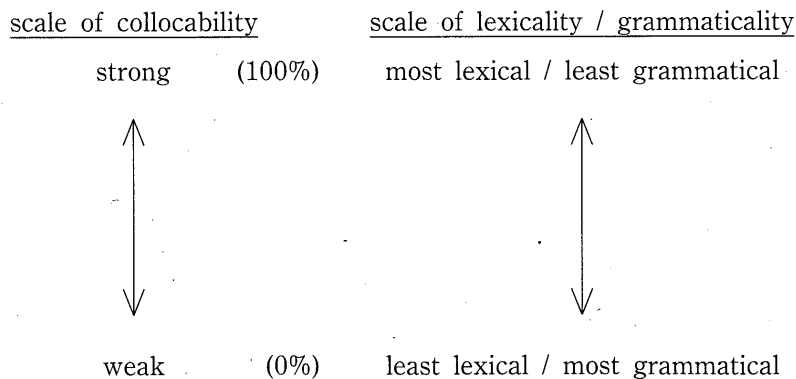
nominal group

e.g. He consented to go.

\*He consented going.

Lexico-grammatical: the word *consent* occurs in adverbial phrases headed by the preposition *by* (e.g. *by mutual consent*) and also collocates with *and* and is a direct object for certain verbs. (e.g. *give consent*, *offer consent*)

It seems that in grammatical collocations, there are no lexical relationship between *city* and *of* in 'the city of Tokyo', and *consent* and *to* in 'He consented to go.' However, Halliday (1966) suggests correlation between 'most grammatical,' 'least lexical' and 'most frequent.' Let us consider the definite article *the* as an extreme case. It is obvious that *the* is most frequently used. Then what do 'most grammatical' and 'least lexical' mean? It does not mean that the definite article *the* does not have any lexical property, but it is best explained grammatically. Let us then take a unique collocator *fro* as the opposite extreme. An item *fro*, which only co-occurs with *to* as in '*to and fro*,' can be regarded as most lexical, least grammatical and least frequent. This means that *fro* has very specific meaning, thus, it can be better understood lexically than grammatically. It seems, then, that we can add another scale, the scale of lexicality alongside the scale of collocability.



## Conclusion

In this paper, Section 1 looked at the theoretical background of collocation referring to the tradition established by Firth, McIntosh, Halliday, and Sinclair and so forth. From those linguists' concepts of collocation, we have established a collocational framework for analysis, which consists of three main areas: strength, typicality, and lexicality, and under each area, types of collocation were examined.

It should be strongly noted that lexis, grammar, and semantics are interrelated. Lexis is not an independent area in the study of language. It is closely interwoven with other areas such as grammar and semantics.

Although this paper did not treat multi-word units / items, the importance of MWU / MWI has been recently pointed out in the study of lexis. (Kadota Ed. 2003) Further study on MWU / MWI should be crucial.

It is considered that in our mental lexicon each lexical item is stored in the network of words. In the process of acquiring a native language, children construct the network of words which are related syntagmatically earlier than the paradigmatic network (Aitchison 2003). This process of vocabulary network development, from syntagmatic to paradigmatic network, applies to language learning (Kadota Ed. 2003). However, how collocation is stored in our mental lexicon has not been elucidated. The investigation of collocation in the mental lexicon seems to be an urgent problem that needs to be studied.

#### References

- Aitchison, J. (2003). *Words in the Mind: An introduction to the mental lexicon*. (3rd ed.). London: Blackwell.
- Carter, R. (1998). *Vocabulary: Applied Linguistic Perspectives* (2nd ed.). London: Routledge.
- Carter, R. and McCarthy, M.J. (1988). *Vocabulary and Language Teaching*. London: Longman.
- Firth, J.R. (1957). *Papers in Linguistics 1934–1951*. London: Oxford University Press.
- Greenbaum, S. (1970). *Verb-Intensifier Collocation in English*. The Hague: Mouton.
- Halliday, M.A.K. (1966). 'Lexis as a Linguistic Level.' In Bazell, C.E. (Ed.) *In Memory of J.R. Firth*. London: Longman.
- Halliday, M.A.K. and Hasan, R. (1976). *Cohesion in English*. London: Longman.
- 池上嘉彦 (編) (1993). 意味論・文体論 (英語学コース 4) 第5版. 大修館書店.
- Jackson, H. (2002). *Lexicography: An Introduction*. London: Routledge.
- 門田修平 (編著) (2003). 英語のメンタルレキシコン 語彙の獲得・処理・学習. 松柏社.
- McCarthy, M.J. (1990). *Vocabulary*. Oxford; Oxford University Press.
- McCarthy, M.J. (2001). *Issues in Applied Linguistics*. Cambridge: Cambridge University Press.
- McIntosh, A. (1966). 'Patterns and ranges.' In McIntosh, A. and Halliday, M.A.K. *Patterns of Language: Papers in General, Descriptive and Applied Linguistics*. 182–199. London: Longman.
- Mitchell, T.F. (1971). 'Linguistic "going-on": Collocations and other lexical matters on the syntagmatic record.' *Archivum Linguisticum* 2: 35–69.
- Nation, I.S.P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Palmer, H.E. (1933). *Second Interim Report on English Collocations*. Tokyo: Kaitakusha.
- Sinclair, J. (1966). 'Beginning the study of Lexis.' In Bazell, C.E. (Ed.) *In Memory of J.R. Firth*. London: Longman.
- Sinclair, J. (et al. Eds.). (1987). *Collins COBUILD English Language Dictionary*. London: Collins ELT.
- 投野由紀夫. (1997). 英語語彙習得論 ポキャプラリー学習を科学する (英語教育リサーチ・デザイン・シリーズ7). 河源社.