

Attaining Fluency in English through Collocations

Han Jinsuk

1. Introduction

Although vocabulary learning is only a part of a language development program, it plays a great role in learning language. From the 1980's communication has been thought of as giving and receiving the content of message, and that we need to acquired as much vocabulary acquisition not only becomes the scale of measuring his/her linguistic competence, but the tools of evaluating the person's knowledge, culture and personality.

Vocabulary acquisition occurs within the exposure of language use through interaction. But in situations like Korea and Japan we don't have an immediate need to use English on social life. We can only be exposed to English through reading books or listening to the radio or manuals. Through reading and listening we come across new vocabulary in context and guess the meaning of words and know the other words. And also we know the syntactic behaviors of the words when they are used and we have a desire to use those words for our own. So we only need various authentic materials to read and through these we can expand our competence in English.

Currently as we focus our attention on the communicative classroom that is directed toward content, tasks or interaction, vocabulary is seen as having a central role in contextualized, meaningful language (Brown 1994;

305). Below are some guide lines for the communicative treatment of vocabulary instruction.

- 1 . As words are basic building blocks of language, survival level of communication can take place quite intelligibly when people simply string words together- without any grammatical rules applying at all. So, if we're interested in being communicative, words are among the first order of business.
- 2 . The best internalization of vocabulary comes from an encounter (comprehension or production) with words within the context of surrounding discourse. Rather than isolating words and/or focusing on dictionary definitions, attend to vocabulary within a communicative frame work in which items appear. Students will then associate new words with a meaningful content to which they apply.
- 3 . We should encourage to develop strategies for determining the meaning of words. A number of "clues" are available to learners to develop "word attack" strategies.
- 4 . Learning of words consumes too much effort and time. Only enduring good language learners can attain successful achievement in vocabulary. Good language learners are creative, developing a "feel" for the language by experimenting with words and grammar, use contextual clues to help them in comprehension, learn to make intelligent guess and learn chunks of language as whole and formalized routines to help them perform beyond their competence. (Rubin Thompson 1982)

1.1. Knowing a word

We need to have stored various kinds of mental lexical knowledge to

comprehend and respond to others and convey the intended message through the function of approach, selection and integration of words. Without a rich vocabulary in mental lexicon, we cannot approach, select and integrate words. Vocabulary is the most important element of achieving proficiency in English. To know a word means the following characteristics. (Carter 1987; 187)

- (1) It means knowing how to use it productively and having the ability to recall it for active use, although for some purpose only passive knowledge is necessary and some words for some users are only ever known passively.
- (2) It means knowing the likelihood of encountering the word in either spoken or written contexts or in both.
- (3) It means knowing the syntactic frames into which the word can be slotted and the underlying forms and derivations which can be made from it.
- (4) It means knowing the relations it contracts with other words in the language and with relations it contracts with other words in the language and with related words in a L1 as well.
- (5) It means perceiving the relative coreness of the word as well as its more marked pragmatic and discoursal function and its style levels.
- (6) It means knowing the different meanings associated with it and, often in a connected way, the range of its collocational patterns.
- (7) It means knowing words as part of or wholly as fixed expressions conveniently memorized to repeat-and adapt- as the occasion arises.

1.2. Basic assumptions

Ultimate goals of language use is to receive and send intended messages

to others fluently and appropriately. To do this we have to know much active vocabulary and this vocabulary should be internalized in mental lexicon in a form of collocation with which language items string together for immediate use without processing. We can assume the following characteristics when we internalized words in collocation:

- (1) Collocation is a network, in language use so if we learn vocabulary, we have to know its collocation.
- (2) Through collocation we know the semantic fields of the vocabulary and their networks. This wide range of knowledge of collocation broadens our linguistic knowledge.
- (3) Language is social convention shared by community. So when we use our language, we use conventionalized expressions to make the hearers understand quickly and to make speaker fluently. So proficiency in language means collocational proficiency.
- (4) For the students to make use of English, they should learn vocabulary in collocation, through which they are accustomed to use the collocation without processing. It is the best way to improve their English Proficiency.

2. How much Vocabulary and how should it be learned ?

Vocabulary knowledge enables language use. Language use enables the increase of vocabulary knowledge (Nation 1993a). Although educated adult native speakers know around 20,000 of word families, a much smaller number of words, say between 2-5,000 word families are needed to provide basis for comprehension. It is possible to make rule of a small number around 2-3,000 words for productive use in speaking and writing.

Clearly a foreign language learner need to know the 3,000 or so high

frequency words of the language. These are an immediate high priority and there is little sense in focusing other vocabulary until these are well learned. The next focus for the teacher is on helping the learners to develop. The next focus for the teacher is on helping the learners to develop strategies to comprehend and learn low frequency words of the language.

A way to manage the learning of huge amount of vocabulary is through indirect or incidental learning (Nation 1990: 16). An example of this is learning new words (or deepening the knowledge of already known words) in context through extensive listening and reading. Learning from context is so important that some studies suggest that first language learners learn most of their vocabulary in this way. (Sternberg 1987). Extensive reading is a good way to enhance word knowledge and get a lot of exposure to most frequent and useful words. At the earlier and intermediate level of language learning, simplified reading books can be of great benefit.

2.1. Vocabulary acquisition from context

People pick up much of their vocabulary knowledge from context, apart from explicit instruction. How much vocabulary growth can be attributed to picking up words from context, and how much is the result of instruction, depends on one's estimate of the total number of words. The importance of context in vocabulary learning is evident from two common sense observations: What a word means on any given occasion is mediated by the many contexts in which it is used, and such contexts provide input from which language users clearly pick up huge amount of vocabulary knowledge, apart from any explicit vocabulary instruction they receive. It seems plausible that first language learners must pick up

most vocabulary from the context, because relatively little of their vocabulary growth can be attributed to vocabulary instruction. Although exposure to a word in a variety of context is extremely important to understanding the depth of the word's meaning provide incidental encounters with words is only one method to facilitate vocabulary acquisition. And also there are some problems using this method. (Sokmen 1997: 237)

- 1 . Acquiring vocabulary mainly through guessing words in context is likely to be a very slow process.
- 2 . Inferring word meaning is an err-prone process. Students seldom guess the correct meaning.
- 3 . Even when students are trained to use flexible reading strategies to guess words in context, their comprehension may still be low due to insufficient vocabulary knowledge.
- 4 . Guessing from context dose not necessarily result in long-term retention.

2.2. Vocabulary Acquisition in collocation and chunks

The term “collocation” is used to refer to group of words that belong together, either because they commonly occur, like “take a chance” or because their meaning is not obvious from the meaning of their parts, as with “by the way” or “to take someone in”(to trick him).

From a learning point of view, it makes sense to regard collocations as items frequently occurring together and with some degrees of semantic unpredictability. This two criteria justify spending time on collocations because of the return in fluency and native-like selection.

- 1) Language knowledge is collocational knowledge. N. Ellis argues

that although it is possible for linguists to discover grammar rules in instances of language, language knowledge and language use can be accounted for by the storage of chunks of language in long term memory and by experience of how likely particular chunks are to occur with other particular chunks, without the need to refer to underlying rules. Language knowledge and use is based on associations between sequentially observed language items. This viewpoint sees collocational knowledge as the essence of language knowledge.

- 2) All fluent and appropriate language use requires collocational knowledge. The best way to explain how language users produce native like sentences and use the language fluently is that in addition to knowing the rules of the language, they store of thousands of preconstructed clauses in their memory and draw on them in language use.
- 3) Many words are used in a limited set of collocations and knowing this is part of what is involved in knowing the words. In some cases the collocations are so idiomatic that they could only be stored as memorized chunks.
- 4) Ellis sees the learning of collocations as one level of “chunking”, that is the development of permanent sets of associative connections in long-term storage.

3. Chunking and collocation

The term “collocation” is used to refer to a group of words that belong together, either because they commonly occur together, like “take a chance”, or because their meaning is not obvious from the meaning of their parts, as with “by the way” or “to take someone in” (to trick them)

A major problem in the study of collocation is determining in a consistent way what should be classified as a collocation. This is a problem

because collocations occur in a variety of general forms and with a variety of relationships between the words that make up the collocation.

From a learning point of view, it makes sense to regard collocations as items frequently occurring together and with some degree of semantic unpredictability. These two criteria justify spending time on collocations because of the return in fluency and nativelike selection.

Collocation is often described as a “Firthian” term (Kjellmer, 1982: 25; Fernando, 1996: 29), but Palmer used it many years earlier and produced a substantial report on English collocations.

Palmer (1933: 4) used a restricted definition of collocation, focusing mainly on items whose meaning is not obvious from their parts:

Each [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 discussed several terms including **idiom**, **heteroseme**, **phrase**, **formula** but decided on **collocation** because it was not a completely new word (Palmer refers to a use in 1750 noted in the Oxford English Dictionary), it had not become definitely associated with other meanings, it was an international word in that it was made of Latin parts, and it could be used in a variety of disciplines.

Considering the role of collocational knowledge in language learning raises an important recurring issue in language study, namely, how much of language learning and language use is based on underlying abstract patterns and how much is based on memorized sequences? When we hear or produce a sentence like “It’s really great to see you !”, do we subconsciously perceive its underlying grammatical structure, do we see it as two or more previously stored chunks “It’s really great” “to see you”, or do we see it as one stored unanalysed chunk that we recognize or

produce when needed? The answer to this question should affect what collocations we give attention to and the way we deal with them in language classrooms. In this chapter we are concerned with collocation but the argument about the units of language knowledge and the way they fit together applies at all levels of language. Let us look first at the units.

3.1. Chunking

In an influential paper, Miller (1956) distinguished “bits” of information from “chunks” of information. Our ability to make reliable one dimensional judgements, such as classifying tones, brightness and size seems to be limited to the same number of items.

“By organising the stimulus input simultaneously into several dimensions and successively into a sequence of chunks, we manage to break (or at least stretch) this informational bottleneck.”(p.95)

Bits of information are formed into chunks by the process of “recording”.

“Since the memory span is a fixed number of chunks, we can increase the number of bits of information that it contains simply by building larger and larger chunks, each chunk containing more information than before.”(p.93)

In his review of one experiment, Miller (1956:94) noted that “apparently the translation from one code to another must be almost automatic or the subject will lose part of the next group while he is trying to remember the translation of the last group.” That is, the recoded items need to be able to be accessed fluently as units in order for them to act as chunks.

Ellis (to appear [Robinson]) sees the learning of collocation as on level of “chunking”, that is, “the development of permanent sets of associative connections in long-term storage” (p.5). This chunking occurs at all

levels of language, and in both spoken and written forms. Table 3.1 has examples from written language.

Leve 1	Type of chunking	Examples
Letter	Each letter is processed as a unit not as a set of separate strokes.	<u>p</u> is processed as a unit, not as a small circle and a descending stroke on the lest hand side
Morphemes	Each morpheme is processed as a unit rather than a set of letters.	<u>play</u> is processed as a unit not as a combination of <u>p</u> , <u>l</u> , <u>a</u> , <u>y</u>
Words	Complex words are processed as a unit rather than several morphemes.	<u>player</u> is processed as a unit not as a combination of two units <u>play</u> and <u>-er</u> .
Collocations	Collocations are processed as a unit not as group of two or more words.	<u>a player with promise</u> is processed as a unit.

Table 3.1 : Examples of chunking and different levels of written language

Chunking can develop in two directions. Memorized unanalysed chunks can be later analysed, or smaller chunks can be grouped into larger chunks. For the moment however let us look at chunking as a process that starts with knowledge of the smallest parts. These small parts are later chunked to become bigger parts and so on.

Chunking typically occurs where the same parts are often observed occurring together. In some cases this occurs solely because of frequency. For example, words like the and soon occur very frequently and may be thus more efficiently treated as one chunk rather than a sequence of letters. In some cases, parts are often observed as occurring together because they represent a regular pattern in the language.

3.2. The advantages and disadvantages of chunking

Then main advantage of chunking is reduced processing time. That is, speed. Instead of having to give close attention to each part, the chunk is seen as a unit which represents a saving in time needed to recognize or produce the item. Instead of having to refer to a rule or pattern to comprehend or produce the chunk, it is treated as a basic existing unit.

Then main disadvantage of chunking is storage. There are many more chunks than there are components of chunks, and if the chunks are also stored in long term memory then there will be a lot of items to store. There may also be difficulty in finding an item in the store.

If chunks are learned as unanalysed units, then a major disadvantage of chunking is that the parts of the unit are not available for creative combination with other parts. For example, if “Please make yourself at home” is learned as an unanalysed unit, then the parts “make yourself...” and “at home” are not available from this chunk to use in other patterns “Make yourself comfortable”, “I really feel at home here” and so on.

The alternative to chunking is rule based processing. In productive language use, this means researching an item each time it is used. The best researched language area on this issue is word building, that is, the use of complex words. When we produce a word like “unable” or “unambiguousness” do we create these words from their parts each time we use them (un+able, un+ambigu+ous+ness) or do we simply retrieve them as already created previously stored complete units? There is a very large amount of research that attempts to answer this important question (see Marslen-Wilson, Kokmisařevsky, Waksler and Older, (1994) for reviews). At present, the research evidence shows that high frequency complex units like “unable” are stored as whole chunks. Low frequency complex items like “unambiguousness” are recreated by rules

each time we need them. If this explanation is correct then it represents a nice compromise between the advantages and disadvantages of chunking. High frequency items are chunked stored separately thus reducing processing time. As we have seen, a small number of high frequency items account for a large proportion of use. Low frequency items are not stored as chunked units, thus reducing the need for lots of storage. As we have seen, there is a very large number of low frequency items which account for a very small proportion of use. This recreation takes processing time but does not happen frequency based balance of storage of chunks and rule based creation or analysis runs through all levels of language.

As chunks become bigger, their frequency of use becomes lower. There will be a point where the frequency of collocations of a certain length is so low that it is not efficient to store them as a chunk. This is a general principle and there will be exceptions where a long collocation is stored as a chunk because an individual uses it frequently. Poems, songs and some speeches are probably also stored in this way.

Type of vocabulary	Number of different words	Coverage of text	Treatment
High frequency words	A few items (not many or store)	A large proportion of text (too much of process)	Store as complete items
Low frequency words	Many items (too many to store)	A small proportion of text (not much to process)	Apply the rules to create them each time they are used

Table 3.2 : Frequency, storage and processing of complex items

This explanation however still does not tell use what the rules are and if there is an interaction between rules and chunks. That is, are rule

based chunks easier to learn? To examine these issues, let us now look in more detail at each of the three positions on collocation that were briefly described at the beginning of this chapter.

3.3. Language knowledge is collocational knowledge

The strongest position taken on the importance of collocational knowledge is that it is essential because the sequential probabilities of language items are the basis of learning, knowledge and use.

In several papers Ellis (in press [Robinson], Ellis and Schmidt, 1997) argues that a lot of language learning can be accounted for by associations between sequentially observation language items. That is, without the need to refer to underlying rules. The major factor affecting this learning by association is frequency of meeting with instances of language use (the power law practice). In essence, holding short sequences of language in short term memory “allows consolidation of long term memory for those same utterances and for regularities or chunks within them... the availability of chunks in long term memory that reflect the regularities in the language means that subsequent utterances that contain these chunks are better perceived and represented in short term memory” (Ellis and Schmidt, 1997: 159). In other words, by having chunks of language in long term memory, language reception and language production is made more effective.

When Ellis talks about “regularities in the language”, he does not mean abstract syntactic frames, but means “frequent sequences, and the more frequent sequences within them” (Eills [in Robinson section 4.1])

3.4 Fluent and appropriate language use requites collocational knowledge

Pawley and Syder (1983) consider that the best explanation of how

language users can choose the most appropriate ways to say things from a large range of possible options (nativelike selection), and can produce language fluently (nativelike fluency) as that units of language of clause length or longer are stored as chunks in the memory. They suggest that this explanation means that most words are stored many times, once as an individual word and numerous times in larger stored chunks.

The “puzzle” of nativelike selection is that by applying grammar rules it is possible to create many grammatically correct ways of saying the same thing. However only a small number of these would sound nativelike. For example, all the following are grammatically correct.

Please close the window.

I desire that the window be closed.

The closing of the window would greatly satisfy me.

The window should be closed please.

Not all are nativelike.

The “puzzle” of nativelike fluency is that we can only encode one clause at a time when speaking and we usually need to do so without hesitations in the middle of the clause.

“Observations of conversational talk indicate that there is a ‘novelty scale’ in the spontaneous speaker’s production of clauses. A minority of spoken clauses are entirely novel creations, in the sense that the combination of lexical items used is new to the speaker; the combination will of course be put together according to familiar grammatical patterns. Some clauses are entirely familiar, memorized sequences. These are strings which the speaker or hearer is capable of consciously assembling or analysing, but which on most occasions of use are recalled as wholes or as automatically chained strings. Still other clauses fall at various points along a cline between these two extremes, consisting partly of new

collocations of lexical items and partly of memorized lexical and structural material” (Pawley and Syder, 1983: 205).

Support for this position comes from a longitudinal study comparing learners of French as a second language before and after residence abroad. Towell, Hawkins and Bazergui (1996) concluded that the observed increase in fluency was the result of proceduralisation of knowledge. This proceduralisation was the result of learners “storing more knowledge as productions in production memory” (p.106). These productions were largely what Pawley and Syder call “memorized sequences”. Towell, Hawkins and Bazergui reached this conclusion by observing that mean length of run (number of successive syllables unbroken by a pause) was the most important temporal variable contributing to the difference between pre-and post-test performance, and by analysing the qualitative changes in some transcripts.

Pawley and Syder argue that “memorized clauses and clause sequences form a high proportion of the fluent stretches of speech heard in everyday conversation” (p.208). Pawley and Syder distinguish “memorized sequences” from “lexicalised sentence stems”. Lexicalised sentence stems are not totally predictable from their parts. They behave as a minimal unit for syntactic purposes, and they are a social institution (a conventional label for a conventional concept). There are degrees of lexicalisation. “Memorized Sequence” are transparent, regularly formed clauses.

Lexicalised sentence stems and memorized sequences are the building blocks of fluent speech. Pawley and Syder (1983: 215) consider that “by far the largest part of a English speaker’s lexicon consists of complex lexical items including several hundreds thousand lexicalised sentence stems.” It is worth stressing that Pawley and Syder are talking about

clause length units, not two or three word phrasal collocations.

To develop fluency, all collocational sequences are important, and they need to be encountered many times, certainly in normal meaning focused use with some pressure or encouragement to perform at a faster speed than a struggling learner usually performs at.

Research on receptive and productive language processing indicates that learners may need experience the language chunks in the medium in which they need to use them. That is, learner are unlikely to become fluent speakers by becoming fluent listeners. To develop speaking fluency they need to practise speaking.

3.5 Some words occur in a limited set collocations

Sinclair (1987) describes two models of the way words occur in a text.

- (1) The *open-choice principles* sees language text as a series of choice where the only limitation on choice is grammaticalness.
- (2) The *idiom principle* sees the constraints and limitations being much greater. As well as limitations based on the nature of the world, and choice of register, “a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (Sinclair, 1991: 110).

The widespread and pervasive nature of the idiom principle is used as a justification for the study of groups of words.

It is not sufficient for this purpose to define a collocation as a group of words frequently occur together. In frequency counts of corpora, the groups although he, but if, and of the frequently occur but do not intuitively fit our idea of what a collocation is. Collocations are closely structur-

ed groups whose parts frequently or uniquely occur together. We would also expect collocations to contain some element of grammatical or lexical unpredictability or inflexibility. It is this unpredictability or learning burden that provides some of the justification for giving collocations special attention in a vocabulary course.

This two part definition of collocation means that groups like eat fish, cold day, and if they would not be considered as collocations, but groups like take medicine, How do you do ?, thin soup would. It is possible to specify further these two general criteria of (1) being closely structured, and (2) containing some element of unpredictability, and later we will look at the scales which can be used to classify and describe collocations.

However, just because a collocation exists does not mean that it deserves attention from a teacher. In order to decide if classroom time and effort should be spent on an item the criteria of frequency and range need to be considered. If the frequency of a collocation is high and it occurs in many different uses of the language, it deserves attention. It must compete for this attention with other collocations and with other words. Frequent collocations deserve attention in the classroom if their frequency is equal or higher than other high frequency words. That is, if the frequency of the collocation would be sufficient to place it in the most frequent 2,000 words, then it clearly deserves classroom time.

Frequent collocations of frequent words also deserve attention. The collocation itself may not be frequent enough to get into the most frequent 2,000 words, but because it is a frequent unpredictable use of a high frequency word, it deserves classroom time. Most collocations deserving classroom time will be of this type, for example, give up, get off, heavy rain.

Let us now look again at the unpredictability aspect of collocations.

The degree of learning burden of a collocation depends on the predictability of its form and meaning. Receptively, as in listening and reading, the learning burden depends on whether the meaning of the collocations is understandable from the meaning of its parts. There are two aspects to this-semantic opaqueness and uniqueness of meaning. The scale of semantic opaqueness involves the degree to which the parts reveal the meaning of the whole. The collocation take medicine can probably be understood from the meaning of its parts with the help of context. Take medicine is not unique in its meaning however as it could mean “consume medicine” or “carry medicine somewhere”

Productively, in speaking and writing, the learning burden of a collocation depends on the predictability of the co-occurrence of its members. Would collocations in the first language or previous learning of the second language allow a user to predict this collocation? Take medicine is not predictable from some learners’ first language

Role of collocational knowledge	Extent	Range of focus	Prototypical activities
Language knowledge= collocational knowledge	Collocational knowledge is the main knowledge	All language items	Unanalysed chunks Dividing up text
Fluent and appropriate use requires chunks	Collocation knowledge is additional knowledge	Many long stretches of items	Fluency activities
Some words have a limited set of collocates	Some words require collocational knowledge	Many words	Study of concordances

(they drink or eat medicine), but the collocations take a pill, take a tablet may be predictable from knowing take medicine.

Table 3.5 : Three positions on the role of collocation

From a vocabulary learning point of view, we need research into collocation

1 . to tell us what the high frequency collocations are

- 2 . to tell us what the unpredictable collocations of high frequency words are
- 3 . to tell us what the common patterns of collocations are where some examples of that pattern would need special attention but where others could be predicted on the basis of this previous attention.
- 4 . to provided dictionaries or information for dictionaries that help learners deal with low frequency collocations.

Knowing the typical collocation of a word is one important aspect of vocabulary knowledge. Firth (1957: 195) noted “part of the meaning of [a] word... can be collocation”, “Meaning by collocation is not at all the same thing as contextual meaning” (p.195). For example, in his description of the collocates of CAUSE are undesirable situations or events like “trouble”, “concern”, “problems,” itself acquires unpleasant connotations and **parole** affects **langue**”.

If collocations are studied because of their unpredictability and frequency of occurrence, we need to know what collocations need an deserve attention. It is this motivation lies behind dictionaries of collocations and frequency studies of collocations.

3.6 Classifying collocations

There is considerable variety in the terms used to describe groups of words which seem to function as units and there are many criteria which are used to classify the groups. The criteria which are used depend on those types of groups that are focused on and the reasons for focusing on them. For example, in focusing on idioms, Fernando (1996) uses the criteria of compositeness (the words fit together as a group), institutionalization (the words frequently occur together), and semantic opacity (the idiom is not the sum of its constituents). Kjellmer (1984) uses six criteria

to measure distinctiveness or degree of lexicalization. These are absolute frequency, relative frequency, length of sequence (number of collocates in the collocation), distribution over texts (range), distribution over text categories (range), and structural complexity. This list is limited because of Kjellmer's aim of using computer-based procedures to find the collocations.

Kennedy (1998: 108–121) notes the wide range of types of collocations and the difficulty in deciding what to classify as collocations. He considers that Firth's (1957: 14) definition of “actual words in habitual company” is central to the definition of collocation but cautions that very large corpora would need to be used to begin to gain reliable and valid data on “habitual” company.

The most effective way of setting up criteria for classifying items as collocations or not collocations and for setting up categories of collocations is to use a set of scales. The large number of scales needed is evidence of the range of items covered by the term *collocation*. We will look at ten scales that have been identified by a variety of researchers. These scales indicate what is involved in learning collocations.

1. *Frequency of co-occurrence*

The most obvious scale ranges from *Frequently occurring together* to *Infrequently occurring together*. In Firth's (1957: 14) words, collocaitons are “actual words in habitual company”. Many studies of collocation exploit this feature by doing computer-based frequency studies of corpora. As we shall see when we look at the other scales, this criterion is not as straightforward as it seems. For example, do the co-occurring items have to be immediately nest to each other, can they change their forms by the use of inflections, do they have to have a strong grammatical

relationship or are common co-occurrences like and the classed as collocations ? Frequency of co-occurrence however is a very important criterion, especially in lists intended for the design of teaching materials. As in studies of vocabulary in the study of collocation, range needs to be considered along with frequency. Kjellmer (1982) provides some instances from the *Brown Corpus* where there are some substantial collocations that occur only in a very limited set of texts.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled, that...

Frequency and range are measured by counting and can be expressed in absolute or relative terms (Kjellmer, 1983: 166-168). Absolute frequency is the actual number of times a collocation occurs in a corpus. Relative frequency compares actual frequency of occurrence compares actual frequency of occurrence with its expected number of occurrences. The expected frequency can be calculated by (frequency of occurrence of item 1 of the collocation) multiplied by (frequency of occurrence of item 2 of the collocation) divided by the size of the corpus. There are other more elaborate ways of taking account of range, and relative frequency.

2. *Adjacency*

Collocates can occur next to each other as in left handed, or separated by variable words or phrases as in little did x realize. The scale ranges from *Next to each other* to *Separated by several items*.

Kennedy (1998) notes co-occurrence of the word silk and a colour such as red, not always necessarily adjacent to each other. Renouf and Sinclair (1991) examine what they call *collocational frameworks* like

be+?+to and too+?+to.

3. *Grammatically connected*

Collocates are usually within the same sentence as a part of a grammatical construction. However it is possible to see items within the same text, not grammatically connected to each other but in a lexical cohesion relationship as collocates. The scale ranges from *Grammatically connected* to *Grammatically unconnected*.

4. *Grammatical uniqueness structured*

Kjellmer (1982: 25) points out that “habitually co-occurring” is inadequate as a criterion because it includes cases like although he, of the and but too, and so it is necessary to have another criterion of grammatical structure. Of the an although he meet the previous criterion of being grammatically connected but they do not make up a collocation that takes account of the major divisions that would be made in analysing a clause. Kjellmer (1982) applied the grammatical structure criterion by using a list of permission structures. The scale ranges from *Well structured* to *Loosely related*.

5. *Grammatically uniqueness*

Some collocations are grammatically unique—hell bent for leather. Others seem to be exceptions to rules—go to bed (bed occurs without an article), and others follow regular patterns weak tea. The scale ranges from *Grammatically unique* to *Grammatically regular* with patterned exceptions like go to bed, town, hospital as the mid-point.

6. *Grammatical fossilization*

Grammatically fossilized collocations do not allow any change to the form of the collocation through a change in word order, for example by and large, law and order or through grammatical change with inflections or part of speech. Some allow small changes: kick the bucket cannot be The bucket was kicked or kicking the bucket, but He kicked the bucket and When do you expect him to kick the bucket ? are possible. Some allow substantial changes in word order: to piece things together can be expressed as thing were pieced together, they were piecing thing together etc. The scale ranges from *No grammatical variation to changes in part of speech, with Inflectional changes as a mid-point.*

7. *Collocational specialization*

Some collocates only occur together. That is, they never or rarely occur without each other, for example Anno-Domini, be-all and end -all, hocus pocus.

Someone collocations consist of one item that only occurs in the presence of the other item, but the other item is not under the same restriction, for example in kith and kin kith seems to be limited to this phrase, while kin can occur in many other places. Other examples include, to and fro, leap year, bubonic plague.

Some collocations consist of items that can also occur with a range of other collocates—good answer, commit suicide. Aisenstadt (1981) calls collocational specialization *restricted connectability*.

The scale ranges from Always mutually co-occurring to All occurring in a range of collocations with one bound item as the mid-point. Renouf and Sinclair (1991) measure the degree of specialization of a collocated by expressing the proportion of times a word occurs in a particular frame-

work as a percentage of the total occurrences of the item. There are now several formula for calculating collocational specialization.

8. *Lexical fossilization*

Some collocations are made up of collocates that cannot be replaced by other words, for examples a bird's eye view, No fear !, by and large. Some collocations allow substitution by words of related meaning, for example entertain a belief, entertain an idea, entertain a desire; last week, last month, last year; last Friday, last Saturday etc. Sinclair (1987) calls this *internal lexical variation*.

The scale ranges from *Unchangeable* to *Allowing substitution in all parts* with *Allowing Substitution in one part* as the mid-point. It is assumed that all substitutions are by semantically-related items, and there is some common meaning in items made on the same collocational frame. The criterion implies that when counting the frequency of collocations the total frequency of the range of permitted substitutions must be counted. See Kennedy (1990) for such an approach to the treatment of preposition-based collocations.

9. *Semantic opacity*

The criterion along with grammatical fossilization are the two most commonly used to define an idiom. The most idiomatic collocations are those where the meaning of the whole is not deducible from the meaning of the parts. Examples include for good, under someone's feet, have a soft spot for someone, of course. The scale ranges from *Semantically opaque* to *Semantically transparent*.

10. *Uniqueness of meaning*

Just as some words have only one meaning, some collocations have only one meaning, for example on be half of, keep secrets, answer the door, full moon. Kick the bucket, however, has two meanings—to die, and to kick the bucket (with your foot). This criterion considers the difficulty learners may have in assigning the appropriate interpretation to a collocation. The scale ranges from *Only one meaning* to *Several Meanings with Related meanings* as the mid-point.

The ranges in each of the ten scales described above have all been graded from most lexicalized to least lexicalized. So, a highly lexicalized collocation would be one like, for example, hocus pocus which is frequent, consists of adjacently occurring items with a strong unique grammatical connection and structure which allows no grammatical and lexical changes, is made up of items that rarely occur individually or in other relationships, whose meaning is not deducible from its parts, and which has only one meaning. Most collocations will be high on the scale for only some of these criteria. The choice, prioritization and weighting of the criteria will depend on the purpose of the classification.

3.7 The Evidence for collocation

There is considerable speculation that collocations are important building blocks in language use and language learning. There are three major types of evidence to support this speculation. First there is the intuitive feeling that certain phrases seem to act as units. Lists of collocations are presented as evidence for this. The work by Pawley and Syder (1983) and Lattinger and De Carrico (1992) is of this type.

Second, there is the evidence from corpus studies that certain groups of words recur. Lists with frequency data are presented as evidence for

this. The work by Kennedy (1992) and Kjellmer (1984) is of this type. This evidence is not easily obtained as collocations are necessarily less frequent than their constituent collocates, and items which intuitively seem to be collocations often have a very low frequency of occurrence in available corpora. Also, evidence using collocations where considerable substitutions is possible does not present a convincing case, because prototypical collocations are those involving very frequent fixed patterns with minimal variation allowed. The more variation and substitution there is in a pattern the more it is towards the grammatical or open-choice end rather than the idiom or lexical end of a collocation scale. Sinclair (1991: 53) argues that “there is a close correlation between the different sense of a word and the structures in which it occurs. “Structures” includes lexical structure in terms of collocations and similar patterns”. Pervasive evidence of this nature provides support for the importance of collocation in language use and language teaching.

Third, there is evidence from studies of learning and knowledge. The work by Towell, Hawkins and Bazergui (1996) is of this type. This kind of evidence shows that language users make use of unanalyzed collocations, that analysed collocations are used with greater speed than would be possible if they were recreated each time they were used, and that there are errors that occur that demonstrate that collocations are being used as lexicalized units. The evidence required is of course the same kind of evidence that is called on in the debate about affixation. That is, are complex work forms like development and developer created from develop plus an affix each time they are used or are they stored for convenience sake as ready-made units? The answer to the question for collocation is likely to be the same as that for the better researched area

of word formation. Some frequent items are treated as lexicalized units, other less frequent items are recreated each time they are used. Items which are frequent and irregular are more likely to be treated as read-made units. One problem with the study of collocation is discovering where the dividing line is.

3.8 Collocation and teaching

To simplify the discussion of teaching, let us consider three points along a scale of collocation. At one end we have idioms like a red herring, you're telling me and be that as it may which are largely fossilized and opaque. In the middle we have groups like take medicine, for example and little did x know which allow some substitution, are sometimes grammatically unique, are not necessarily adjacent and are at least partially transparent. At the other end we have items like as a result, it is assumed that, where was I ? which are grammatically well formed, allow a lot of substitution and grammatical change, and are transparent.

Idioms need to be dealt with as if they were words in that they should be given attention on the basis of their frequency and range of occurrence. Learning their meaning should be enriched by analysis and explanation of their parts and history, and some attention should be given to the way they function in discourse.

The items like take medicine which are to some degree unpredictable (some languages say drink medicine, some others eat medicine) need to be examined for any patterning that occurs (take medicine, take a rest, take a break, take a holiday). Very frequent collocations can be the starting point for dealing with the range of related collocates.

The very predictable collocations should be dealt with as part of the enrichment of the individual collocates that make them up. For exam-

ple, when the learners meet a word like *clear*, they should be introduced to its more common collocates such as a clear day, a clear sky, a clear thinker, a clear road. Some very frequent or immediately useful collocations like Can you tell me where the toilet is, please ? can simply be memorized and used, and later be analysed when the learner's level of proficiency is more advanced.

The principle of learning burden applies just as much to collocations as it does to individual words. The learning burden of an item is high if its form, meaning and use are not readily predictable from previous first language or second language knowledge. Its learning burden is light if it follows regular predictable patterns. There are numerous patterns of regularity lying behind groups of collocations.

3.9 Encouraging chunking

Chunking can develop from known parts. It can also occur from the memorization of unanalyzed chunks. There are two major approaches to help learners chunk known compartments. The most important is to help them in larger units. It is likely that this fluency development is to some degree skill specific so that learners would need to have fluency practice in listening, speaking, reading, and writing. Ellis and Laporte (in press) note that “at least for beginning learner, there are strong benefits of output practice in both the SLA of vocabulary and of phrases and collocations.”

The fluency strand of a course is an essential components in the development of chunking. We will briefly overview activities for the development of fluency in each of the four skills shortly.

The second major approach to help learners chunk is through deliberate language focused attention. This attention can involve practice in chun-

king text containing familiar items, and the deliberate teaching and learning of collocates of known items. This can include the use of concordances, matching activities, and the development of collocation tables.

The memorization of unanalyzed chunks is an important learning strategy, especially for a learner who wants to quickly gain a degree of fluency in limited areas. It has other learning benefits as well, particularly in that it quickly provides a fund of familiar items that can be later analysed to help support the development of rules.

3.10 Chunking through fluency development

Schmidt (1992) presents a comprehensive survey of a wide range of theories which can be used to explain fluency development. The most accessible theory that describes the development of chunking through fluency development is McLaughlin's (1990) restructuring theory, McLaughlin (1990: 113) argues that there structuring of language knowledge, which for our purposes we will see as chunking, occurs when learners reach a high degree of automatization.

“... practice can lead to improvement in performance as sub-skills become automated, with attendant decrements in performance as learners reorganize their internal representational framework. In the second case, performance may follow a U-shaped curve, declining as more complex internal representations replace less complex ones, and increasing again as skill becomes expertise.”

This means that learners can become fluent through practice at one level of knowledge. the only way they can improve further is to restructure that knowledge, perhaps into larger chunks. This will slow them down initially, but they will then be able to reach higher levels of fluency

because of restructuring. McLaughlin these sees fluency development playing a central role in chunking.

When examining activities to see if they are likely to help the development of fluency, a teacher should look for the following features.

- 1 The activity should involve only known vocabulary and grammatical features, and preferably should involve familiar content knowledge. This can be achieved by working with material that has already been studied in previous classes, by choosing very simple material, by allowing learners to control the task, and by helping learners to plan.
- 2 The activity should be meaning focused. That is, the learners should be interested in and focused on the messages they are sending or receiving.
- 3 There should be some encouragement to do the activity at a speed that is faster than the learners' normal speed. This should be possible because the learners are working with familiar simple material. The encouragement can be in the form of time pressure, competition (with one's own previous performance or with others), or repetition.
- 4 The activity should involve a large quantity of language processing. That is learners should be reading or writing texts several hundred words long, or speaking and listening for several minutes.

Here are some activities that meet these requirements and can thus be considered to be fluency development activities.

The 4/3/2 technique was devised by Maurice (1983). In this technique, learners work in pairs with one acting as the speaker and the other as listener. the speaker talks for four minutes on a topic while her partner listens. Then the pairs change with each speaker giving the same infor-

mation to a new partner in three minutes, followed by a further change and a two-minute talk.

A **Listening corner** is a place where the learners can listen to tapes as a part of self-access activities. The teacher makes a tape of a spoken version of writing that the learners have already done. The writing could be done individually or as group compositions. Instead of learner compositions learners can listen to recordings of what they have read before (in English or the first language), such as the reading texts from earlier sections of the course book.

Listening to stories is particularly suitable for learners who read well but whose listening skills are poor. The teacher chooses an interesting story possibly a graded reader and reads aloud a chapter each day to the learners. The learners just listen to the story and enjoy it. While reading the story the teacher sits next to the blackboard and writes any words that the learners might not recognize in their spoken form. Any words the learners have not met before may also be written, but the story should be chosen so that there are very few of these. During the reading of the first chapters the teacher may go fairly slowly and repeat some sentences. As the learners become more familiar with the story the speed increases and the repetitions decrease. Learner interest in this activity is very high and the daily story is usually looked forward to with the same excitement people have in television serials. If the pauses are a little bit longer than usual in telling the story, this allows learners to consider what has just been heard and to anticipate what may come next. It allows learners to listen to language at normal speed without becoming lost. The graded readers In the Beginning (Longman Structural Readers, Stage 2), Of Mice and Men (Heinemann Guided Readers, Upper level) and Animal Farm (Longman Bridge Series) are particularly good.

The best recording is a useful fluency activity involving a tape recorder or the language laboratory. The learner speaks on to the tape talking about previous experience or describing a picture or set of pictures. The learner listens to the recording noting any points where improvement could be made. Then the learner re-records the talk. This continues until the learner is happy with the recording. This technique can involve planning and encourages repetition through the setting of a quality-based goal.

Rehearsed talks involve using the pyramid procedure of preparing a talk individually, rehearsing it with a partner, practising it in a small group, and then presenting it to the whole class.

Speed reading and **Extensive reading** of graded readers provide fluency improvement through the features of limited demands because of language control, and quantity of processing. To be effective, speed reading courses need to be written within a limited vocabulary so that learners can focus on the reading skill without having to tackle language difficulties. Speed reading courses also have the added benefit of involving the learners in keeping a running record of their speed and comprehension scores. Research on graded readers (Wodinsky and Nation, 1988) shows that reading only a few books at one level would provide the learners with contact with almost all the words at that level. This shows that graded reading can provide a reliable basis for systematic coverage of vocabulary for fluency development.

Repeated reading is one approach to developing fluency in reading (Dowhower, 1989; Rasinski, 1989). The learners read the same text several times. There are several ways of doing this. One way is to set the learners a new task to do each time so that each reading is for a different purpose. The tasks would become more demanding with each

repetition. Another way it to set a time goal for reading the text, say, 3 minutes for a 500 word text. The learners reread the text until they can do it in the set time. An even simpler goal is to get the learners to reread the text a set number of times. Research suggests that 4 or 5 times is most effective (Dowhower, 1989).

Continuous writing is an activity where learners are given a set time (usually 5–10 minutes) to write with the aim of producing a large quantity of writing within the time. The learners can record the number of words they wrote on a graph. the teacher responds to the writing not by correcting errors but by finding something positive in the content of the writing to comment on briefly.

These fluency development activities cover a range of skills and apply the our criteria of familiarity, meaning focus, pressure, and quantity.

3.11 Memorizing unanalysed chunks

A very useful strategy, particularly in the early stages of language learning, is to memorize useful unanalysed chunks. This strategy can be applied to both regularly formed and irregularly formed chunks. The regularly formed chunks may eventually be analysed and form the basis for learning grammatical patterns.

Chunks can be most effectively memorized by applying the same learning guidelines as for isolated words. These are

- 1 Write each chunk on a small card with its translation on the other side so that there has to be active retrieval of its form or meaning.
- 2 Repeat the chunk aloud while memorizing it.
- 3 Space the repetitions so that there is an increasingly greater interval between learning sessions.

- 4 Use mnemonic tricks like the keyword technique, putting the chunk in a sentence, visualizing examples of the meaning of the chunk, and analysing the parts of the chunk. This increases the quality of the mental processing and helps learning.
- 5 Don't learn chunks with similar words or meaning together. They will interfere with each other.
- 6 Keep changing the order of the cards to avoid serial learning.

Teachers need to develop an awareness of the difficulties that lie behind some collocations and the kinds of patterns that exist. This will enrich their teaching and allow them to focus their effort of productive patterns where possible. Teachers also need to know why attention to collocation is useful and know a well balanced range of ways of giving this attention.

4. Conclusion

As communication is sending and receiving messages we need to store as much vocabulary as possible. To make communication fluent and appropriate we have to use collocation automatically. Automatic use of chunks of words is like the large building blocks of a house. Difficulties of processing words. Therefore the use of chunks of collocation makes English use more fluent and appropriate. The knowledge of collocation is one's competence of English.

Works cited

- Brown. H. D. (1994) *Teaching by Principles*, Prentice Hall Regents
Principles of Language Learning and Teaching, 3rd Edition, Prentice
Hall Regents

- Carter. R (1987) *Vocabulary*, Allen & Unwin
- Carter. R & McCarthy. M (1988) *Vocabulary and Language Teaching*, Longman
- Hoey. M (1996) *Patterns of Lexis in Text*, Oxford Univ. press
- Fernando.c (1996) *Idioms and Idiomaticity*. Oxford Univ. press
- Firth. J. R (1957) *Papers in Linguistics* (1934–1951), Oxford University press
- Kennedy. G (1990) *Collocation; where grammar and vocabulary Teaching meet*, Singapore
- Kjellmer. G (1990) *Pattern of Collocability*, in *Corpus Linguistics*
- Maurice (1983) *The Fluency workshop*, *TESOL Newsletter* 8; 83
- McCarthy (1989) *Some Vocabulary pattern in conversation*
- Nation. I. S. P (1999) *Learning Vocabulary in Another Language*, Victoria Univ. of Wellington
- (1990) *Teaching and Learning Vocabulary*, Newbury House
- Nunan. D (1999) *Second Language Teaching & Learning*, Heinle & Heinle
- Read. J (2000) *Assessing Vocabulary*, Cambridge Univ. press
- Taylor. L (1990) *Teaching and Learning Vocabulary*, Prentice Hall International.