A Study of Common Language Use as Evidenced by the British National Corpus (BNC) and its Implications for Learning and Teaching

Aisha Attaher Orayith

a.orayith@yahoo.com

Abstract

The aim of this study is to use the spoken section of BNC to identify commonly used collocations of spoken English in the belief that these frequently employed, and therefore useful items should be prioritised for teaching to elementary level learners in the EFL classroom. The Compleat Lexical Tutor was used as a tool to access the BNC spoken section as it provides extensive range of samples of British spoken English. It was found that many casual collocations are frequently used in spoken English; these were compiled into a list. Elementary and beginner course books were then analysed in order to investigate the extent to which collocations are used in the early stages of learning. The study offers some examples of collocations and ideas for presenting and integrating them into the curriculum, before presenting the implications of the findings for learning and teaching.

Key words:
collocations, corpus, corpora, concordance, BNC.

Received 29/11/2020
Accepted 16/12/2020
Published online 19/12/2020

1. Introduction

The aim of this study is to use the spoken part of the British National Corpus (BNC) to identify high frequency collocations of spoken English and to use these findings to prioritise the most frequently used, and therefore potentially most useful, language for teaching in the EFL classroom. The study discusses the differences between collocations and other word groups before ultimately presenting a list of commonly used collocations for low level learners of English. These were identified by applying a strict set of criteria. Current course books were analysed in light of the findings with a view to identifying how they might be improved.

1.1. Background to corpus linguistics

McEnery and Hardie (2011) define corpus linguistics (plural corpora) as the study of language data on a large scale through the computer-aided analysis of collections of written texts or transcribed utterances. There are many types of corpora, including general, specialised and learner corpora. According to Hunston (2002), well-known general corpora include the British National Corpus (BNC) of 100 million words and the Bank of English of 400 million words. McEnery and Hardie (2011) argue that corpus linguistics is an area that allows the researcher to investigate various aspects of language by examining how it is used. In other words, it can be employed to study not just lexis but also other language systems such as grammar. Concordancing is the main method used to study corpora.

This study employs the BNC as it is the best-known national corpus in Britain and it is generally considered to represent modern British English. It adopts concordancing as the main method to find the answers to the research questions.

1.2. BNC (British National Corpus)

The British National Corpus (BNC) is the central research tool in this study. It is described as a monolingual, general corpus covering Modern British English of the late twentieth century. Its samples cover a wide range of styles, varieties, genres and registers. Written samples account for 90 per cent of the corpus and the remaining 10 per cent of the corpus are spoken texts. The corpus can be used for a range of linguistic research, such as the study of lexis, morphology, syntax and semantics and for discourse analysis. It can also be used for English language teaching purposes such as the design of materials and syllabuses, for classroom usage and for independent study (BNC Webmaster, 2009).

1.3. Research questions

This study seeks to answer the following four sub-questions:

- What are the most frequent collocations of spoken English?
- What are the most common collocational patterns?
- What criteria are needed to distinguish collocations from other word groups?
- What are the implications of the study findings for language learning and teaching?

2. Literature Review

Previous studies, scholastic articles and published research will be presented and critically discussed to provide insight into and understanding of the various current theories and approaches towards collocations in English language teaching.

2.1. What is collocation?

Collocation can be defined in several ways. *The Cambridge Online Dictionary* (2020), for example, gives more than one definition for collocations, or collocates. One of these definitions is:

"A word or phrase that is often used with another word or phrase, in a way that sounds correct to people who have spoken the language all their lives, but might not be expected from the meaning" (p.1).

They give the example *a hard frost* to show that it can be difficult for a language learner to guess a collocation's meaning. The learner might use the adjective *strong* instead, but this will not sound natural or true. Another definition, provided by Yamasaki (2008), is that a collocation is the appearance of two, three or four words near each other in a text, or "the habitual co-occurrence of individual lexical items" (Crystal, 2003, p.82).

The importance of collocations is increasingly widely recognised, with several studies demonstrating that knowledge of collocations is a defining marker of the achievement of fluency and parity with native speakers (Chon and Shin, 2013; Nesselhauf,2005). Thornbury (2005,p.23) echoes this argument when he asserts that: "speakers achieve fluency through the use of prefabricated chunks".

2.2. Distinguishing collocations from other word groups

One of the ways of distinguishing collocations from free combinations is by the level of frequency with which they appear. This way of distinguishing collocations is called the frequency-based approach, according to Nesselhauf (2005). There are two additional criteria that can be used to distinguish collocations from other word groups: opacity and commutability, also known as transparency and combinability respectively (Nesselhauf, 2005). Trantescu (2010) explains that opaque collocations carry metaphorical meanings and are often restricted to a specific domain or

situational context. Collocations are often perceived as being semantically transparent, but they in fact have varying degrees of semantic transparency. It is also often assumed that their elements in combination carry the same meaning as when used separately, but this is not necessarily the case. The same author explains that a collocation is transparent when all elements of the collocation add an aspect of their transparent meaning. Nesselhauf (2005) argues that the distinction between collocations and other word groups is most often made on the basis of commutability only, because it is generally thought to be more relevant and it is also easier to measure than opacity.

2.3. Recent corpus-based studies on collocation

The study on which this research is based was that conducted by Shin and Nation (2008). These authors used the BNC to identify a list of the most common collocations in spoken English, arguing that these might usefully be taught in elementary speaking courses. The researchers reported identifying 4,698 collocations. The most significant finding was that many of these qualify for the inclusion in the most common 2,000 words of English, if collocations were not distinguished from single words. In terms of method, Shin and Nation (2008) used the spoken section of the BNC as their source of data. They used WordSmith Tools 3.0 to search for all of the occurrences of the pivot word and create concordances, which were then checked manually. The study produced four main findings:

1. Many grammatically well-formed collocations occur among the most common collocations.

2. The more common the pivot word, the more collocations were found. It was proved that the first 100 pivot words have a larger number of collocations than the second 100 and so on.

3. An extremely large proportion of collocations are made up of a very small number of pivot words.

4. They also concluded that: "The shorter the collocation, the greater the frequency" (p.344). They found that 77 per cent of the collocations are two-word combinations.

Other previous studies

Chin & Shin (2013) also conducted a very similar study in which they identified academic collocations in spoken and written discourse in the British Academic Spoken English and the Academic Corpus. They studied 20 target words in each corpus, which produced 934 written and 460 spoken collocations. The study made a comparison of the top 50 collocations which suggested raising awareness of common collocations in the field of economics.

In his study, Kennedy (2003) explored the BNC in order to find possible collocations of 24 amplifiers. The analysis intended to demonstrate how corpus-based analysis can be used to understand the structure of English collocations. Implications for language teaching were considered too. Statistical analysis was applied by calculating the frequency of co-occurrence of 24 amplifiers (8 maximizers & 16 boosters). The study produced lists of the most common 40 collocations of each adverb. It was found that not all amplifiers can be interchangeable and some adverbs collocate distinctively from the others. It was also found that some verbs and adjectives tend to collocate more strongly with some particular adverbs. For instance, *fully & perfectly* collocate with adjectives of positive associations only, such as *fully informed & perfectly acceptable*. However, *totally & utterly* have negative associations, such as *totally illogical & utterly useless*. Whereas, *entirely* has both positive and negative associations.

2.4.Corpus impact on language learning and teaching

Corpus linguistics has had a number of positive effects on English language teaching and learning. O'Keefee *et al.* (2007) and Kovacs (2013) all argue that dictionary-making has benefited significantly, with dictionaries such as *Macmillan Collocations Dictionary* (2010) being based on language corpora. These dictionaries have in turn influenced language teaching materials. Language corpora have also made it possible for users to check their intuitions about language and find out whether their ideas are correct by conducting empirical studies of language use. Harris and Jaén (2010) point out that teachers and learners alike can access several online corpora for free.

Harmer (2013) agrees with O'Keefee *et al.* (2007) that corpus linguistics can be useful for ELT and makes the additional point that lexicographers can now be more confident when making statements about lexis because they are able to analyse language using data-driven programs. O'Keefee *et al.* (2007) observe that language

corpora can be used for computer assisted language learning (CALL) by giving students tasks to do or materials based on the corpus. Harris and Jaén (2010) agree that corpus consultation can be integrated into CALL.

3. Methodology

3.1. Qualitative and quantitative research approaches

Baker (2013) states that corpus linguistics is considered a quantitative research tool, though McEnery and Wilson (1996) argue that corpus analysis can be either quantitative or qualitative; it can be quantitative if the research involves classifying or counting features or constructing statistical results, but it can be qualitative if the research does not involve assigning frequencies but aims mainly for a detailed description. The research questions in this study necessitate quantitative analysis of the BNC as they involved counting language patterns and assigning frequencies. However, the qualitative approach will also be employed in the use of content analysis to investigate the implications for learning and teaching.

3.2. The corpus as a research tool

Baker (cited in Littoselitti, 2013) states that the corpus is a research method rather than a branch of linguistics like semantics or pragmatics, while Kheirzadeh and Marandi (2014) say it can be a useful source of learning materials that can provide data about different parts of the language. There are two main approaches to researching corpus linguistics: the corpus-based approach and the corpus-driven approach. The difference is that the former usually uses the corpus to examine already existing theories that are based on other data sources, while the latter usually uses the corpus to develop original linguistics hypotheses (ibid). This study is corpus-based in that it uses the BNC as a method to test an already existing theory about the most common collocations in spoken English.

3.3. The computer program: the Compleat Lexical Tutor

Through doing considerable research, it has been found that numerous programs can be used to search the corpus, including the WordSmith tools, AntConc, the Compleat Lexical Tutor and the BNC web. For this study**,** the Compleat Lexical Tutor will be used to explore the collocational relationships of the pivot words. Such programs are able to search the corpus quickly and easily, though not all results will be useful; for

example, one might get a chunk that appear repeatedly or syntactically incomplete collocations (O'Keefee*et al*. 2007).

3.4. Concordance

According to Wray and Bloomer (2006), the concordance program allows the researcher to discover what sorts of words usually occur in the immediate context of a certain word . O'Keefee et al. (2007) describe concordancing as one of the basic techniques when using a corpus. Harmer (2013) provides a simple definition for concordancing or as he puts it "concordancers". He defines it as: "a selection of lines from the various texts in the corpus showing the search word in use" (p.34). Concordancers can be used to search for word collocations through the use of computer software like WordSmith Tools or Montconc Pro to find every occurrence of a target word, which is known as the node, are displayed in lines in a format where the node is in the centre of the line with the accompanying context to the left and right sides. This is known as a Key Word In Context display or KWIC concordance. The concordancing tool employed in the current study to search for collocations of the keywords in context was the Compleat Lexical Tutor.

3.5 Criteria used for identification and coding of collocations

The adopted criteria were similar to those used by Shin and Nation (2008), though fewer pivot words will be used as the scale of this study is smaller. Only ten target words will be used as it was assumed that these ten words would retrieve an appropriate number of collocations for teaching to elementary level learners. Using a larger number of target words would have retrieved far too many collocations for teaching at this level. According to Shin and Nation (2007), the criteria for selecting target words are:

1. Each target word should be a different word type.

2. Each target word should be a content word.

3. All ten target words should occur among the most frequent 100 words on Leech, Rayson and Wilson's (2001) list.

4. The collocation should occur at least 30 times in 10 million running words.

5. Collocations must not cross an immediate constituent boundary.

6. If a collocation has more than one meaning, each meaning should be considered as a separate collocation.

3.6. Selection of the pivot words

Collocations associated with ten content words (listed in Table 1) will be the focus of this study. These ten pivot words were selected because they occur among the 100 most common words in the BNC corpus, as listed by Leech *et al*. (2001), and because they meet the criteria listed in Section 3.5. The frequency cut-off point of a word is that a string must occur at least ten times per million words (Cortes, 2002; O'Keefee*et al*.,2007). However, the Leech *et al*. (2001) list also adheres to a minimum of ten occurrences per million words.

Word	**PoS**	**FrS**
Is	Verb	10164
Do	Verb	9594
Was	Verb	8097
Have	Verb	7488
Be	Verb	5790
Know	Verb	5550
Well	Adverb	5310
So	Adverb	5067
Got	Verb	5025
Are	Verb	4663

Table 1: 10 pivot words

Source: Leech, Rayson and Wilson (2001)

4. Results and Discussion

The study findings will be presented to answer the first three research questions; the results pertaining to the final question (implications for language learning and teaching) are presented in the next section.

BNC study findings

4.1. What are the most frequent collocations of spoken English?

The collocation	Frequency/ 1 million	Frequency/ 10 million
you know	1002	10021
to do	1001	10005
do you	1000	10004
have to	1000	10003
to be	1000	10002
it is	998	9981
is it	791	9706
don't know	904	9045
this is	802	8024
is that	773	7733

Table 2: The 10 most frequent collocations

The study retrieved 219 results and this list addresses the first research question. Table 2 above shows the ten most common collocations. Some are verb – noun agreements, for instance *it is*, *it was* and *do you*. Two of them can be classified as to-infinitive: *to be* and *to do*. The most significant result is that the combination *you know* tops the list, occurring 1002 times in a million running words. This gives an interesting insight into the interactional nature of spoken English.

4.2.The most useful collocations

The collocation	No. per 1 million	No. per 10 million
You know	1002	10021
I don't know	702	7021
As well	637	6373
Gonna be	193	1931
Got it	180	1804
So much	134	1335
And so on	87	871
There you are	55	552
The thing is	51	508
Well done	44	435
Is about	42	420
I don't think so	37	367
As well as	62	619
***You was**	30	302
Be careful	25	247
So sorry	5	46

Table 3: The most useful collocations

Table 3 shows word combinations that occur frequently in spoken English; learning these at elementary level will improve learners' speaking skills by making their English sound more natural. As stated in 4.1, the collocation *you know* is useful because of its interactional nature. *There you are* and *here you are* also have an interactional function as they are used when handing something to someone – these are arguably useful phrases for the target learners. And *so on* is another interactional

collocation that can be very useful to cut a long list short and to show that something is repetitive or the same as information already given. *The thing is* is commonly used in spoken English as an introductory phrase, making it useful to elementary level learners. They should also be introduced to *got it*, and that it can mean "I understand you, I received the message you are trying to communicate".

Although not high up in the list, *you was* is an important combination and teachers should alert students to it. Although it is non-standard, it is commonly used by English native speakers when speaking. I would therefore argue that learners should at least be able to comprehend it when they hear it, even if they cannot use it, to facilitate understanding and prevent confusion. Similarly, learners also need to know *gonna do*, *was gonna* and *gonna be* to be able to understand what people are saying to them, though they may not be ready to use these combinations themselves at this level. Again, *gonna* is very frequent in the British corpus, despite being non-standard and it is thought to be a colloquialism derived from American English.

Furthermore, students can also be introduced to the combination *is about* and its meaning that the activity is going to be done in the very near future. For example,"The train is about to leave" means that the train will close the doors and leave in seconds. In this context, the collocation might be extremely useful as it could save the learner from missing the train. The collocation *I don't know* will also benefit elementary level learners if presented as meta-language that they can use in the classroom.

The following collocations are also thought to be very useful, although less frequent: *well then, the problem is, the important thing is, that'll do, was born, to be honest, be careful, as you know, let me know, well known* and *so sorry*.

4.3.What are the most common collocational patterns?

Examination of the retrieved list revealed that the most common grammatical categories are:

1) Pronoun+ verb, for example *you know*, *it is*, *you was* and *I think so*

2) Modal+ bare infinitive, for example *must have*, *should be* and *might have*

3) Adverb+ adverb, for example *so far*, *so much* and *very well*

4) Verb+ pronoun, for example, *was it, do yo*u and *got it*

Most of the results retrieved are two-word combinations. However, there are 39 collocations of three-word combinations and six collocations of four-word combinations. I would therefore argue that short collocations are the most common.

4.4.The number of grammatically well-formed high frequency collocations

The hypothesis that there are a large number of grammatically well-formed collocations is supported by the list. Most, if not all, the results are well-formed grammatical phrases, as noted in Section 4.3. However, it is worth mentioning the collocation *you was* as it lacks verb to noun agreement, which makes it ungrammatical. *You were* is a grammatically well-formed combination, but many native English speakers tend to use "was" with "you" in their speech. Learners should be aware of this and accept both combinations but know when to use each form. Other ungrammatical items are the collocates of *gonna*, for instance *is gonna* and *gonna do*, as mentioned earlier.

4.5.Comparison between the frequency of the pivot words and the number of their collocates.

Word	FrS	No. of collocates
Is	10164	55
Do	9594	34
Was	8097	18
Have	7488	19
Be	5790	28
Know	5550	27
Well	5310	11
So	5067	14
Got	5025	18
Are	4663	12

Table 4: Comparison between the frequency of the pivot words and the number of their collocates

One can argue that the more common the pivot word, the more collocates it is likely to retrieve. The most common pivot word (*is*) retrieved 55 collocates, which is more than the number of collocates (34) retrieved by the second most common pivot word (*do*). At the other end of the list, the least frequent pivot word (*are*) retrieved only 12 results. However, this was not the smallest number of collocates – the pivot word *well* retrieved only 11 collocates. By examining the other pivot words and comparing their results, it was noted that a couple of pivot words retrieved more results than others which are more frequent. A case in point, the pivot *was* is more frequent than *have*, but it retrieved fewer collocates; the same applies to *so* and *got*.

4.6.Comparison between the length of the collocation and its frequency rate

The shorter the collocation, the more frequent it is. This is evident in Table 2, which shows that the most common collocations are two-word combinations. Nine of the ten collocations in the table are two-word; the odd one out was *don't know,* which can be considered a two-word combination because of the contractions that are always applied in spoken English. In fact, there are 124 two-word combinations, which make up 56 per cent of the total number of the retrieved collocations. Three-word combinations are much less frequent; for example, *there you are* only occurs 55 times per million words. Four-word collocations are even less frequent than this; the collocation *I don't think so* occurs only 37 times per million words. Extending the analysis to the top 100 collocations reveals that 93 per cent are two-word collocations. On the other hand, analysis of the bottom 100 collocations reveals that the majority are long. It can therefore be concluded that the longer the collocation is, the less frequent it is.

4.7.Frequency cut-off point of single words and the most common collocations

Collocations	**99**	**77**	**22**	**1**	**22**
Words	1st 1000	2nd 1000	3rd 1000	4th 1000	5th 1000
Cut-off point	76/1 million	32/1 million	19/1 million	13/1 million	

Table 5: Frequency comparison between single word types and collocations

For a single word type to be included in the top 1000 words of the spoken corpus, it should have a frequency of 760 times per 10 million or 76 occurrences per 1 million. Table 5 shows that 99 collocations meet the cut-off point of 76 occurrences per million words. Consequently, these collocations can be considered alongside the first 1000 word types. Correspondingly, 77 collocations could be included in the top 2000 category (frequency rate higher than 32occurrences per 1 million). The 99 collocations in the first band include *I don't know*, *you know* and *and so on*. The 77 collocations in the second band include *there you are*, *as well as* and *well, then*.

Thus, if no distinction is made between collocations and single words, many collocations meet the frequency criterion and can therefore be included in the most frequent 1000 and the most frequent 2000 words in English. However, there are fewer collocations in the remaining categories.

4.8.What criteria are needed to distinguish collocations from other word groups?

The aim of this study is to search for significant collocations. However, none were retrieved. The results retrieved are all casual collocations, though some are arguably useful to learn at elementary level. Casual collocations are those in which both elements can be used with any other word, while significant collocations are those in which one of the elements can only occur with the other element. Thus, casual and significant collocations are differentiated in terms of their commutability. The same concept, in conjunction with the frequency rate of their common recurring patterns (30 per 10 million words), has also been used to distinguish collocations from other word groups. Some of the results retrieved in this study are commonly used word combinations (e.g. *this is, he was* and *we have*), while others are casual collocations (e.g.*so sorry*, *well done* and *as you know*).

4.9. Analysis of course book content

4.9.1. Use of collocations in the elementary course book

The analysis of *New English File Elementary* showed that most of the vocabulary lessons deal with single words on topics such as food, clothing and family members. However, the book does introduce students to collocations of common verbs. Units 3.b, 4.a and 5.c deal with verb phrases such as collocates of *go*, *have* and *get*. The teacher's book suggests that the teacher should encourage the learners to use the

whole phrase; that is, the verb and its collocate. Another interesting issue to raise here is the use of boxes to introduce learners to common phrases, for example *here you are*, *let's go* and *come on.* It is worth noting that most of these phrases are used in spoken English, not in written English. As the collocations retrieved in this study were not found in the elementary level book, the beginner level book was analysed in order to find out whether these collocations are dealt with at an earlier stage.

4.9.2. Use of collocations in the beginner course book

Most of the vocabulary lessons in *New English File Beginner* deal with single words and topics similar to those in the elementary level book, such as numbers, food and family members. However, units 3.a, 4.a, 4.b and 5.c did present sets of phrases with common verbs, for instance *go home*, *have a shower* and *get up.* The teacher's book highlights that the teacher should raise students' awareness of verb-noun collocations, for example *play tennis* and *watch TV* and it also encourages the teacher to show the learners how to differentiate between *make a meal*, *do the housework* and *do homework*. There are boxes with useful daily phrases to learn. These phrases are quite useful – it was noted that the collocation *I don't know* is taught via one of the boxes. This collocation occurred 702 times in a million running words. The boxes also present the common collocation *it is going to* in complete sentences. This study found that the collocation *is going to* occurred 103 times in a million running words. However, although the colloquial *gonna* occurred frequently in the results in combination with *was*, *do* and *have*, it is not introduced in the course book.

5. Implications

the implications of the study findings are drawn out in order to answer the final research question. The findings of the study, in regard to this and the other research questions, are then compared to previous literature.

5.1. What implications do the study's findings have for language learning and teaching?

This section answers the final research question by discussing the ways in which the study's findings might influence language learning and teaching. As Section 4.9 shows, the course books include some collocations which though not of high frequency, are very useful for everyday life. The teacher could adapt the course books

by replacing, adapting or adding activities, such as gap fill activities or word maps, to include the most common collocations. They should also include the very frequent collocations (see Section 4.1) in the syllabus in addition to raising students' awareness of the importance and usefulness of the phrases that are included in the course book.

Furthermore, learning spoken collocations is arguably more important to ESOL learners. As they need to use English in their daily lives, they need to understand features of the spoken language – they need to be familiar with the communicative exchanges that will allow them to take part in conversations in an effective way (Watkins, 2005). In contrast, EFL learners might need to learn or focus more on written English than spoken English.

Given the importance of collocations in spoken English, lessons and course books focusing on spoken language should give them particular emphasis. This can be achieved in different ways, some of which are outlined in the following sections.

5.1.1. Integration of collocations into language study boxes

High frequency collocations can be integrated into the English language teaching curriculum through the use of language study boxes. The *New English File* course books use boxes to introduce learners to useful phrases. similar boxes can be incorporated to introduce learners to the most common collocations. For example, the interactional collocations could be grouped together in one box and presented along with speaking practice or when the lesson is focusing on spoken language. Ungrammatical and non-standard collocations could also be grouped together in one box and presented along with the relevant grammar point. A vivid case in point, the collocation *you was* could be introduced when teaching the past tense of the copula verb *be*.

5.1.2. Integration of collocations into word maps or word circles

Using a word map, the teacher can show learners how words group together (Harmer, 2013). Thus, the map can be used to show possible collocates of a word that the learners have learnt recently. The teacher can prepare the word map for the learners, or they can use it as an activity in which the learners try to add as many collocates as possible to the map.

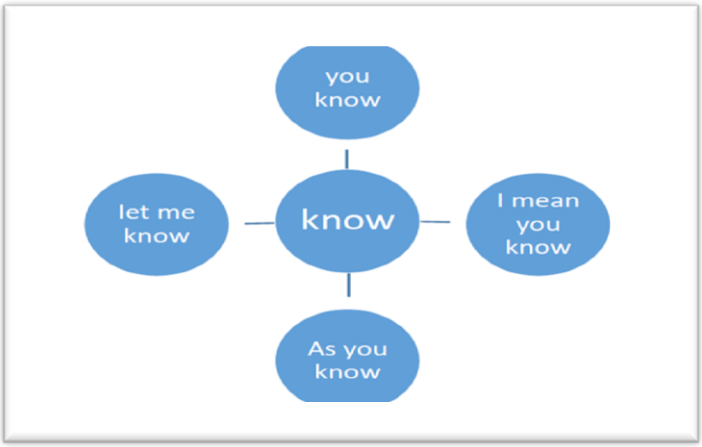

Figure 1: word map of collocates of 'know'

Another useful alternative is a word circle. Harmer (2013) explains that learners can be asked to decide which words from the wheel can be combined with the main word and which words cannot go together and are not possible combinations.

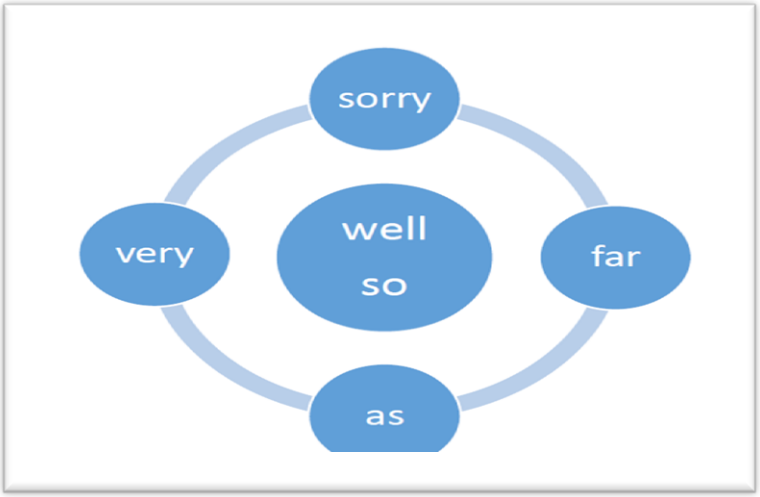

Figure 2: word circle of 'well' & 'so'

Williams (2014) argues that learners should get used to recording collocations in different ways, including word maps, word circles and boxes. His observation that the process of language learning is not linear but cyclical reminds me of the importance of recycling vocabulary, which these activities can also be used to do.

5.1.3. Implicit vs. explicit syllabus

The content of the language learning syllabus is central to successful language learning. It dictates lesson planning and the teaching process (Kennedy, 2003). The

collocations retrieved in this study do not all need to be explicitly taught as some are examples of colloquial language that elementary students do not need. However, if fluency is the teacher's aim, these commonly used collocations could be embedded in the teacher's own speech. Kennedy (2003) explains that they can be part of an implicit curriculum, through which the teacher allows the learners to acquire certain language items through exposure. He argues that it is essential for learners to acquire "experience of which linguistic items typically occur in the company of other items" (p.483). Language experts and materials designers may not agree with the inclusion of casual collocations in the curriculum, but teachers can nevertheless include some of these in fluency activities, according to their appropriateness and usefulness for learners.

It is worth mentioning here that knowing the collocations of a word is one aspect of knowing that word. Gass and Mackey (2013) argue that vocabulary acquisition can be placed on a continuum that starts with the learner having a passive (receptive) knowledge of a word– possibly only a superficial familiarity – and ends with them having an active (productive) knowledge that sees them able to use the word accurately. Pignot-Shahov (2012, p.38) distinguishes between the two ends of the continuum thus: "Productive knowledge is usually associated with speaking and writing, while receptive knowledge is associated with listening and reading". Elementary level learners should at least acquire passive (receptive) knowledge of these collocations and become familiar with their use and meaning to avoid confusion and facilitate interaction.

5.1.4. Maximising internalisation opportunities

Kennedy (2003, p.483) argues that: "The challenge for language teachers is to devise a curriculum that maximises the opportunities for learners to get enough experience of the units of language in use in order to internalise them." The first step towards achieving this is to ensure that learners are frequently exposed to the prefabricated units in a meaningful context. The extent to which collocations are learnt largely depends on the frequency with which learners meet these word combinations. However, although exposure can help improve learners' knowledge, it may not be enough to teach them to use collocations themselves (Nesselhauf, 2005; Bhans,

1993); in other words, it may not suffice to take them from receptive to productive knowledge.

Lightbown and Spada (2013) argue that it can be useful for learners to imitate language chunks. They observe that recent corpus studies have identified an increase in the use of formulaic language, and that research has also shown that language is learnt in chunks (ibid). It might be argued then that it should be taught in chunks. Learning language in chunks can also reduce the number of errors that result from transferring and translating from a learner's first language. The technique of backward build-up (expansion) drill that is used in the audio-lingual method of language teaching can be useful for learning collocations as it can be employed to help learners acquire language chunks rather than single words.

5.1.5. Raising awareness of learners and teachers

It is very important to raise teachers' awareness of the commonly used collocations, their common patterns and the criteria needed to distinguish them. As Kennedy (2003, p.484) argues: "The most important outcome of corpus-based insights into what language learning entails may be in consciousness-raising for teachers". Language teachers should make an effort to use these collocations in the classroom either in their own speech or in fluency activities, for example by creating scripted role plays that feature collocations. The learners' attention should be drawn to the meanings and uses of these combinations through demonstration, pictures or in meaningful situations. Shin and Nation (2007) argue that care should be taken when the teacher is choosing which items to use from the list as they are derived from a spoken corpus and some of them are colloquial. Consequently, it is better if they are presented and identified in connected speech, such as recorded authentic conversations.

5.2. Comparison with the adapted study and other studies

There are some similarities and differences between this study's findings and those of Shin and Nation (2007).The main difference is the number of collocations retrieved; Shin and Nation (2007) retrieved 4,698, while this study retrieved 219 collocations. As in this study, only ten pivot words were used in contrast to Shin and Nation's (2007) 1000 pivot words. It was thought that ten pivot words would retrieve enough collocations for elementary level learners in light of Shin and Nation's (2007) finding that the first 100 pivot words have an average of20.5 collocations each. This was

confirmed when this study identified an average of 21.9 collocations for each of the ten pivot words.

The results for the first research question (refer to Section 4.1) showed some points of similarity between the two studies with the inclusion in the list of *you know*, *as well*, *so much* and *very well*, though the rest of the collocations are completely different. It should be borne in mind, however, that Shin and Nation (2008) only presented the first 100 collocations in their results. In other words, it is not possible to make a definite statement that the items identified in this study were not among Shin and Nation's results since they did not publish their entire list. In both of these studies, and in that by O'Keefee *et al*. (2008), the collocation *you know* has a high frequency score. The actual frequency rate differs significantly, however; in this study, this item has a frequency rate of 10021 per 10 million words, while in Shin and Nation's study it has a frequency rate of 27348 per million words. Both studies identify it as the most common collocation in the BNC spoken corpus. Thus, I would argue that it should be taught to elementary level learners as it will be helpful for them when interacting and communicating in English. The collocation *very well* also appears among the results of Shin (2007), where it has a frequency rate of 48. In this study, it scored a frequency rate of 80.

In essence, the interactional nature of the collocations retrieved from both studies is apparent, giving the reader an interesting insight into the nature of spoken English. This type of collocation should be expected as the results were retrieved from a spoken corpus, and not from a written or academic corpus. Furthermore, both studies agree that short collocations are used more frequently than longer ones. In Shin and Nation's (2008) study, two-word collocations make up 77 per cent of their results. Similarly, two-word collocations make up 56 per cent of the total number of collocations in the results of this study. Additionally, in both studies the top 100 collocations are mostly two-word collocations, while the majority of the bottom 100 collocations are long. Interestingly, Ellis *et al*. (2008) found that three-, four- and five-word sequences are more common in academic than in non-academic English. However, Shin (2007) found that two-word collocations are more common in both spoken and written English.

A further important principle is the number of collocations meeting the cut-off point to be included in the first, second, third, fourth and fifth 1000 words in English. The 1000 word level was reached by 99 collocations in this study but by only 84 collocations in Shin and Nation's study, though their study used more pivot words and thus retrieved more results. On the other hand, only 77 collocations met the cut-off point to be included in the second 1000 words, while 224 collocations did this in the adapted study. Following on from this point, the cut-off point for the first 3000 words was met by more than 500 collocations in Shin and Nation's study. However, only 22 met the cut-off point in this study.

Following on from the previous point, one can therefore conclude that the more common the pivot word is, the more common collocations it will retrieve. This study used the ten most common words in spoken English as pivot words and retrieved 219 collocations. A large number of these collocations met the cut-off level of frequency to be included in the first 1000 words in English. In contrast, Shin and Nation's (2007) study used the first 1000 pivot words and retrieved 5,894 collocations, but only a small number of these collocations achieved this level of frequency.

In answer to the second research question, it was found that both studies retrieved many grammatically well-formed high frequency collocations. The collocations retrieved in this study were categorised according to their grammatical components; these are different from Kovacs's (2013) categories.

As regards the third research question – what criteria are needed to distinguish collocations from other word groups – Shin and Nation's study did not provide an explicit answer to this question, rendering comparison impossible. However, the findings here can be compared to those of other prominent writers such as Nesselhauf (2005) and Menon and Mukundan (2012), who also support the use of commutability- and frequency-based approaches as a way of distinguishing collocations from other word groups. The frequency-based approach can distinguish collocations from other less common word combinations, while commutability, or combinability, distinguishes casual collocations from significant collocations and the latter from idioms through the degree of substitution of its elements. Other authors, such as Laufer and Waldman (2011), argue for distinguishing collocations on the basis of verbs and nouns.

5.3. Conclusion and limitations

To sum up, the BNC was used as the research method to create a list of the most common collocations in English. This corpus was selected because it provides numerous samples of British spoken English. Examination of the corpus revealed that there are many casual collocations that are frequently used in spoken English.

There were some limitations to the present study, the main one being the time limit, which affected the number of pivot words it was possible to investigate. It is hoped that further research can add to this study. Future research could study more pivot words; 100 words would probably be sufficient to retrieve more useful and possibly significant collocations. Future teachers and researchers may gain insight from this research as it presents reliable information and demonstrates a procedure that can be applied to other corpora.

References:

BNC Webmaster. 2010. *Corpus*. available at: http://www.natcorp.ox.ac.uk/corpus/index.xml (accessed 02/11/2014)

Budden, J. 2010. *Recycling vocabulary*. Teaching english. British council. Retrieved from: http://www.teachingenglish.org.uk/language-assistant/teaching-tips/recycling-vocabulary (accessed 27/3/2015)

Cambridge Dictionaries Online. 2014. Cambridge University Press. retrieved from:

http://dictionary.cambridge.org/dictionary/british/collocation (accessed 19/11/2014)

Chon, Y. V. & Shin, D. 2013. A corpus-driven analysis of spoken andwritten academic collocations. *Multimedia-Assisted Language Learning, V.16*(3), pp.11-38.

Cortes, V. 2002. Lexical Bundels in Freshman Composition, in Biber, D. Fitzmaurice, S. and Rippen, R. (eds) Using Corpora to Explore Linguistics Variation. Philadelphia. P.A.Jhon Benjamin, pp. 131-145.

Crystal, D. 2003. *A Dictionary of Linguistics & Phonetics.* Fifth Edition. Oxford: Blackwell Publishing.

Davies, G. 2015. CALL (computer assisted language learning). LLAS Centre for Languages, Linguistics and Area Studies, University of Southampton. Retrieved from: https://www.llas.ac.uk/resources/gpg/61 (accessed 06/05/2015)

Dörnyei, Z. 2007. *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford ,OUP.

Gass, S.M. and Mackey, A. 2013. The Routledge Handbook of Second Language Acquisition. London. Routledge.

Harmer, J. 2013. *The Practice of English Language Teaching*. 4th ed. Essex. Pearson Education Limited.

Harris, T, & Moreno, JM (eds) 2010, Linguistic Insights, Volume 128 : Corpus Linguistics in Language Teaching, Peter Lang AG, Berne, IN, USA. Available from: ProQuest ebrary. [23/11/2014].

Hunston, S. 2002. *Corpora in Applied Linguistics*. [Online]. Cambridge Applied Linguistics. Cambridge: Cambridge University Press.

Jens Bahns. 1993. Lexical collocations: a contrastive view. *ELT Journal*. V.47(1) January 1993, pp.56-63

Kennedy,G. 2003. Amplifier Collocations in the British National Corpus: Implications for English Language Teaching. *TESOL Quarterly*, V. 37(3).pp. 467-487. Teachers of English to Speakers of Other Languages, Inc. (TESOL)

Kovács, É.2013.On Collocations in English. *British and American Studies*, V. 19. P. 201-209.

Larsen-Freeman, D. and Anderson, M. 2011. *Techniques and Principles in Language Teaching*. 3rd ed. Oxford; Oxford University Press.

Leech, G., P. Rayson, and A. Wilson. 2001. *Word Frequencies inWritten and Spoken English: Based on the British National Corpus*. London: Longman.

Lewis, M. 2000. "Learning in the lexical approach" in *Teaching Collocation: Further Developments in the Lexical Approach*. Michael Lewis (ed.). England: Language Teaching Publications, pp. 155-158.

Lightbown, P. and Spada, N. 2013. *How Languages are Learned*. 4th ed. Oxford; Oxford University Press.

Litosseliti, L.2010. Research Methods in Linguistics. London. Continuum international publishing group.

Longman dictionary of applied linguistics, retrieved from: http://site.ebrary.com/lib/nottinghamtrent/reader.action?docID=10074867 (accessed 09/11/2014)

Mart, C. 2013. The Grammar-Translation Method and the Use of Translation to Facilitate Learning in ESL Classes. Journal of Advances in English Language Teaching. V1(4), pp.103-105

McEnery, T and Hardie, A. 2011.Corpus Linguistics: Method, Theory and Practice. New York. Cambridge University Press.

MenoN, S. and Mukundan, J. 2012. Collocations of High Frequency Noun Keywords in Prescribed Science Textbooks. International Education Studies. V.5(6). Canadian Center of Science and Education.

Nesselhauf, N.2005. Collocations in a learner corpus. Philadelphia: John Benjamins.

O'Keeffe, A. McCarthy, M. and Carter, R.2007. From Corpus to Classroom: Language Use and Language Teaching. Cambridge. Cambridge University Press,

Oxenden, C. , Latham-Koeing, C. And Seligson, P. 2004. New English File Elementary. Oxford University Press.

Oxenden, C. and Latham-Koeing, C. 2009. New English File Beginner. Oxford University Press.

Oxford online dictionary. 2014. Oxford University Press.

Pignot-Shahov, V. 2012. Measuring L2 Receptive and ProductiveVocabulary Knowledge. LANGUAGE STUDIES WORKING PAPERS. Vol.4, pp.37.45 University of Reading.

Shin, D. 2007. The High Frequency Collocations of Spoken and Written English.English Teaching, V.62(1).

Shin, D. and nation, P.2008. Beyond single words: the most frequent collocations in spoken English. ELT Journal, V.62(4). Oxford University Press

Thornbury, S. 2005. How to Teach English. Pearson Education Limited. Edinburg.

TRANTESCU, A.M. 2010. Semantic transparency and opacity in fixed expressions.Annals of the University of Craiova. Series Philology. Linguistics.

Watkins, P. 2012. Learning to teach English. Surrey. Delta Publishing.

Williams, B. 2014. Collocations with advanced levels 2- classroom activities. British council.

Yamasaki,N.2008. Collocations and colligations associated with discourse functions of unspecific anaphoric nouns. International Journal of Corpus Linguistics, V. 13(1).pp.75–98.John Benjamins Publishing Company

Your voice (2014). BBC: www.bbc.co.uk/voices/yourvoice/conversation2.shtml