

## A Comparative Study of Right Collocates of *Tiny*, *Small*, and *Minute* in the British National Corpus

Nouman Hamid<sup>1\*</sup>, Sehrish Shafi<sup>2</sup>

### Abstract

The meaning of a word can be established, in part, by looking at the words frequently collocating with it. The present study establishes and compares the various senses and meanings of three apparently synonymous adjectives - *tiny*, *small*, and *minute* - by examining their immediate right collocates. The study draws upon the British National Corpus (BNC), comprising 100 million words. The results reveal that the three selected adjectives differ not only in their frequencies and register distribution but also in their sense and meaning. Out of the three adjectives, *small* is most frequently used in the BNC. Moreover, the positive forms of all three adjectives are most commonly used in the BNC as compared to their comparative and superlative forms. As regards the meaning, although there are some similarities, each of the three adjectives conveys certain sense(s) which are not shared by the others. The study concludes that the three selected adjectives are not strict synonyms and cannot be used interchangeably in all contexts. The findings of this corpus-based investigation are also compared with the definitions and illustrations of the three selected adjectives in *Oxford Advanced Learner's Dictionary* (2010). The results of the present study have implications for lexicography and English language teaching.

**Keywords:** Collocation, Collocate, Semantic Sense, Adjective, Lexicography

### Introduction

*Collocation* is one of the most controversial concepts in linguistics, although it is based on widely shared intuition that certain words have a tendency to occur near each other in natural language, for example, *ring* and *bell*, *kick* and *bucket*, etc. (Johansson & van Waarden, 2024; Evert, 2007). However, the term has been defined and used differently by various linguists belonging to different schools of thought in the past five

<sup>1</sup> Department of English, University of Azad Jammu and Kashmir, Muzaffarabad

<sup>2</sup> Department of English, Mirpur University of Science and Technology, Mirpur

\*Corresponding author's E-mail: [noumanhamid@gmail.com](mailto:noumanhamid@gmail.com)

Available online: 01-03-2025

This is an open-access article.

to six decades. In fact, it has been approached from three prominent standpoints.

First, the viewpoint from which it was not only defined by Firth (1957) but many linguists of the present era is what is known by the *empirical view*. The proponents of the *empirical view* used this term for characteristic and frequently recurrent word combinations, arguing that the meaning and usage of a word can to some extent be characterized by its most typical collocates (Kim et al., 2024; Evert, 2007).

Second, the term was also defined in the field of phraseology to refer to semi-compositional and lexically determined word combinations like *stiff drink*, *heavy smoker*, etc. In this approach, collocates are divided into subcategories ranging from purely opaque idioms to semantically compositional word combinations, which are merely subject to arbitrary lexical restrictions (Evert, 2007, p. 2). According to this view, semantic relation between collocates is more basic than the syntactic one. Another important characteristic of collocation, as highlighted by this view, is the semi-compositionality, which means that the meaning conveyed by a pair of collocates is not necessarily the sum of the meaning of its parts (individual words).

Although both empirical and phraseological views agree on classifying most word pairs as collocates, they differ in examples where collocates give compositional meaning. For instance, a pair like *bad* and *time* is considered a collocate in the empirical view, but it is not held as such in the phraseological view (Coffey, 2022).

Finally, in computational linguistics, collocate is a generic term for any lexicalized word combination that has idiosyncratic semantic or syntactic properties and may therefore require special treatment in a machine-readable dictionary or natural language processing system (Evert, 2007, p. 3). In the present study, following Evert (2007), McEnery and Wilson (2001), and McEnery, Xiao, and Tono (2006), the terms *collocation* and *collocate(s)* are used in the empirical sense.

Firth (1957) introduced the term *collocation* when he investigated the frequent collocates of *ass*. He finds that there are very few adjectives which can collocate with *ass* (*silly*, *obstinate*, *stupid*, *awful*, and *young*) (Krishnamurthy, 2000) and it gives one of its meanings when it is immediately preceded by *you silly* or other forms of address. Similarly, he states that one of the meanings of *night* is attributed to its tendency to collocate with *dark* (Finlayson et al., 2024; Greaves & Warren, 2010).

Halliday (1966) was the first linguist who felt the need to measure the distance between two collocating items in a text. He also brought in the concept of probability in the research on collocations and, following empirical view, emphasized the need for data, quantitative analyses, and the use of statistical measures instead of simple frequency information (Krishnamurthy, 2000).

Before the development and popularity of machine-readable corpora, lexicographers and linguists used to undertake the daunting task of analyzing collocates manually. Sinclair (1970) was probably the first linguist who developed computational methods of looking at collocations in a corpus and brought in the parameter of position, which meant that collocations of very frequent words were positionally restricted (Kopotev, 2024; Krishnamurthy, 2000).

Evert (2007) presents the most frequent collocates of *bucket* in the BNC using different association scores (MI and simple-II), which take into account observed counts (O), frequency of first word ( $f_1$ ), frequency of second word ( $f_2$ ), and total number of words in a corpus (N). Using the threshold level of  $f \geq 3$  and word span of L5 R5, he found that the most frequent collocates of *bucket*, on the basis of Simple-II association measure, were *water, a, spade, plastic, size*, etc., whereas the ones on the basis of MI were *fourteen-record, ten-record, full-track*, etc. (*bucket* used in the technical sense as a data structure in computer science). It shows one of the flaws of MI measure, which will be discussed later (Pu et al., 2024).

With the help of corpus data, Moon (2010) shows that verbs of motion are usually followed by adverbials or prepositional phrases of direction or manner. He also reports that the verb *comply* is usually preceded by something that indicates coercion, necessity, willingness, etc. (*incentive, must, force, fail(ure), hesitate*), and/or is followed by *with*, itself typically followed by a noun phrase indicating a constraint (*agreement, decision, law, obligation*, etc.) (p. 200).

O’Keeffe *et al.* (2007) find that the verbs *go* and *turn* are similar in the sense that both collocate with *grey, brown, and white*, but they are different due to the reason that they do not always both collocate with many other words. For instance, *go* can combine with *mad, insane, bald, and blind* but *turn* cannot collocate with these (Greaves & Warren, 2010).

Biber, Conrad, and Reppen (1998) compare the immediate right collocates of *big, large, and great* in academic prose and fiction registers of Longman-Lancaster corpus. They find that *big* frequently collocates

with *enough* and *traders* in academic prose, and with *man*, *enough*, *and*, *black*, *house*, etc. in fiction. *Large*, on the other hand, frequently collocates with *number*, *numbers*, *scale*, etc. in academic prose and with *and*, *black*, *enough*, *house*, *room*, etc. in fiction. *Great* frequently collocates with *deal*, *importance*, *number*, *majority*, etc. in academic prose and with *deal*, *man*, *burrow*, *big*, etc. in fiction (Aldereihim, 2023).

Comparing different meanings and senses conveyed by *big*, *large*, and *great* in connection with their most frequent collocates, Biber, Conrad, and Reppen (1998) observe that in both academic prose and fiction, *big* is most commonly used to describe the physical size of objects. On the other hand, the most frequent collocates of *large* in academic prose indicate that it most commonly refers to a quantity or amount of something, physical size, and magnitude of various processes; whereas it is commonly used to describe physical size in fiction. It is less commonly used to refer to an amount or quantity in fiction. The adjective *great*, apart from referring to an amount or quantity, is used to show intensity in academic prose which makes it different from the other two adjectives; while in fiction, it refers to amount, a sense of importance, and, less commonly, large physical size. They argue that the three adjectives differ in meanings due to different words with which they frequently collocate (Riches et al., 2023).

Research on collocations has benefitted lexicography and dictionary writing a great deal. McEnery, Xiao, and Tono (2006) compare the immediate right collocates of adjective *sweet* in the first (intuition-based) and fourth (corpus-based) edition of the *Longman Dictionary of Contemporary English* (henceforth LDOCE1 for first edition and LDOCE4 for fourth edition) considering the most frequent collocates of *sweet* in the BNC. They observe that the use of corpus data has brought improvements in LDOCE4 in two ways. First, LDOCE4 defines and illustrates *sweet* in a much greater detail and with the help of a large number of examples than its definition, illustration, and number of examples in LDOCE1. Second, more collocates (42.86 percent) of *sweet* appearing in the example sentences provided in LDOCE4 fall in the top ten most frequent collocates of *sweet* in the BNC than the ones provided in LDOCE1 (33.33 percent). This shows that paying attention to most frequent collocates has helped improve dictionary entries (González-Díaz, 2021).

Based on the most frequent left and right collocates of *deal* as a noun in a sample from London-Lancaster corpus, Biber, Conrad, and Reppen (1998) highlight the most important uses of *deal* as a noun and compare

their findings with the definitions of *deal* provided in five different dictionaries. They observe that dictionaries include one or more of the following seven senses of *deal* as a noun: a) a large but indefinite amount, b) an agreement or arrangement, c) the distribution of cards in a game, d) treatment received, e) the act of distributing, f) wood of fir or pine trees, and g) a business transaction (p. 39). Comparing these seven senses of *deal* as a noun in five dictionaries, they conclude that although most dictionaries cover all seven senses, they differ in the order in which these senses are presented. Comparing these seven senses with their own findings, they point out that the corpus-based analysis yielded additional senses of *deal* as a noun which were not covered in many dictionaries. Moreover, one sense of *deal*, mentioned in many dictionaries (distribution of cards in a game) was not found in the corpus-based analysis (Dang et al., 2022).

Although there is a fair amount of published research on collocation in general and on collocates of adjectives in particular, the researchers were unable to locate any research focusing on the comparison of the three selected adjectives in terms of their most frequent collocates and the similarities and differences in their meanings due to their collocates. Moreover, there is an apparent semantic similarity among the three adjectives selected for this study which makes English language learners assume that these are synonymous and can be used interchangeably but this assumption has never been tested with empirical evidence. To bridge this gap, the present study seeks to compare the immediate right collocates of the three seemingly synonymous adjectives (*tiny*, *small*, and *minute*) in the British National Corpus to establish and find out the similarities and differences among the senses and meanings of these adjectives and to compare the findings with their definitions given in *Oxford Advanced Learner's Dictionary* (2010). The present study seeks to answer the following research questions.

- What are the most frequent immediate right collocates of *tiny*, *small*, and *minute* in the British National Corpus?
- What are the semantic similarities and differences among *tiny*, *small*, and *minute* with respect to their most frequent collocates?
- How do the findings compare with the definition(s) of these adjectives in *Oxford Advanced Learner's Dictionary* (2010)?

## Methodology

### Data Source

The British National Corpus (BNC) (accessed through Brigham Young University's English-corpora.org) was used as the data source because it is a) publicly available, b) tagged for parts of speech, and c) considered as an authentic and representative sample of British English. It consists of approximately 100 million words and is one of the largest corpora of the English language. 90 percent of the corpus consists of different types of written texts, whereas 10 percent is dedicated to the transcribed version of spoken texts. Written section of the BNC includes texts from a variety of disciplines including natural science, applied science, social science, world affairs, commerce, arts, etc., while spoken texts represent different situations such as business, institutional, leisure, conversations, lectures, etc. The BNC was constructed between 1960 and 1993. A number of variables were taken into consideration while including texts in the BNC such as gender, age, education, etc. (McEnery, Xian, & Tone, 2006; Meyer, 2004). The corpus is divided into seven sub-sections (registers), which include spoken, fiction, magazine, newspaper, non-academic, academic, and miscellaneous. Out of these registers, miscellaneous and magazine contain most and least number of words, respectively. Table 1 provides the total number of words for each register included in the BNC.

**Table 1.** *Number of words in sub-sections of the BNC*

No.	Sub-section	Number of words (approximate)
1	Spoken	10 million
2	Fiction	16 million
3	Magazine	7 million
4	Newspaper	10 million
5	Non-academic	16 million
6	Academic	15 million
7	Miscellaneous	20 million

### Type of Co-occurrence

Evert (2007) identifies three types of co-occurrences (collocates), which include a) surface co-occurrences (two words which appear within a certain distance measured by the number of intervening word tokens); b) textual co-occurrences (two words which appear in the same textual unit, that is, sentence, utterance, or even the whole document); and c) syntactic

co-occurrence (two words which have a direct syntactic relation between them, e.g., part of the same noun phrase, etc.). Of these, the present study focuses on surface co-occurrences.

### **Statistical Measures**

Several statistical measures are used to determine the strength of association between members or parts or constituents of collocates. The most common of these is raw frequency which shows the number of times two words co-occur in a corpus or sub-corpus. However, it does not take into account the frequencies of individual words, high frequency words (grammatical words) tend to be the most frequent collocates of a node.

Z score is another statistical measure used for the strength of collocation. It compares the observed frequency with the frequency expected if only chance is affecting the distribution (McEnery, Xiao, & Tono, 2006, p. 57). The higher the z score, the greater the degree of collocability. However, this statistical measure assumes data to be normally distributed, which is not true of most corpus-based text analyses. Due to this reason, it lists many rare words as the top collocates of a node. That is why, this is less frequently used in corpus-based studies of collocation (Evert, 2007; McEnery, Xiao, & Tono, 2006).

Yet another statistical measure of association is t score which is computed by subtracting the expected frequency from the observed frequency and then dividing the result by the standard deviation. A t score of 2 or above is normally considered to be statistically significant, though the specific probability level can be looked up in a table of distribution, using the computed t score and the number of degrees of freedom (McEnery, Xiao, & Tono, 2006, pp. 56-57).

Finally, there is MI (Mutual Information) score, which takes into consideration observed frequency as well as expected frequency (computed by multiplying the raw frequencies of first and second word in a collocate and dividing the result by total number of words in the corpus) (Evert, 2007). It measures the possibility of two words appearing together within a specified span of words (Biber, Conrad, & Reppen, 1998). A higher MI score is associated with a strong link between two words. If MI score is close to 0, it means that the two words co-occur most probably by chance. A negative MI score indicates that the two items tend to shun each other. An MI score of 3 or higher is considered to be evidence that the two items are collocates. However, in practical applications, MI is found to have a tendency to assign inflated scores to

low-frequency word pairs, especially for data from large corpora. Thus, even a single co-occurrence of two-word types might result in a fairly-high association score (Biber, Conrad, & Reppen, 1998; Evert, 2007; McEnery & Wilson, 2001; McEnery, Xiao, & Tono, 2006).

Many other statistical measures are also used by researchers to determine the strength of association between a node and its collocate, including log-likelihood (LL) score, MI3 score, log-log score, etc. (McEnery, Xiao, & Tono, 2006). However, only raw frequency and Mutual Information (MI) score are used to generate the most frequent collocates of *tiny*, *small*, and *minute* in the present study since these scores can be automatically computed by the BNC. Other scores, although more robust, cannot be easily computed. Moreover, the two measures (raw frequency and MI scores) were also used to find out the similarities and differences in the most frequent collocates of the three selected adjectives generated by the two measures.

### **Procedures for Data Analysis**

First, the raw frequency lists were generated for the three selected adjectives on the basis of lemma which were manually converted into normalized frequencies. Second, the three adjectives were compared based on their raw and normalized frequencies. Third, lists of ten most frequently occurring immediate right collocates were generated for the three adjectives. Word span was set to L0 R1. Fourth, each list was re-organized on the basis of MI scores in descending order. Fifth, based on the raw frequencies, top ten collocates of each adjective were analyzed to find out its meanings. Sixth, the meanings derived for the three adjectives were compared to find out the semantic similarities and differences among them. Finally, the findings were compared to the definitions of the three selected adjectives given in *Oxford Advanced Learner's Dictionary* (2010).

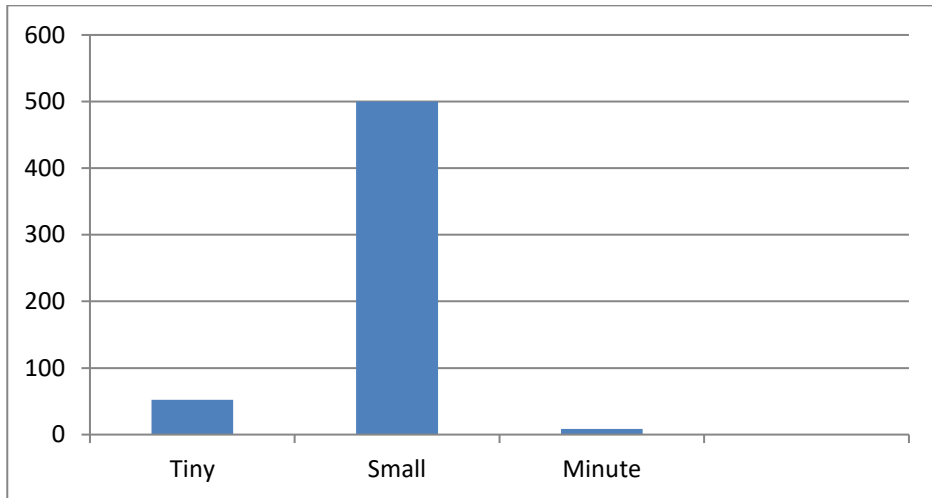
## **Results and Discussion**

### **Overall Frequencies of *Tiny*, *Small*, and *Minute***

An open lemma search for the three selected adjectives provided instances of their use not only as adjectives but also as nouns (surprisingly, *small* and *tiny* were also found to have been used as nouns). In order to avoid this, more restricted search strings were used. For instance, in order to generate the raw frequency of the lemma *tiny*, [tiny].[j\*] was used as the search string which resulted in generating the raw frequency of the lemma *tiny* used as an adjective in the BNC. Figure



1 presents the comparison of the three adjectives based on their normalized frequencies.



**Figure 1** *Normalized Frequencies of Tiny, Small, and Minute*

Figure 1 shows that the lemma *small* was found to be the most frequent among the three adjectives with a normalized frequency of 499.78 per million words, followed by *tiny* with 52.36 per million words, which was followed by *minute* with 8.42 per million words.

### Frequency of the Lemma *Tiny*

A glance at the different forms of the lemma *tiny* in the BNC, as presented in Table 2, reveals that the positive form (*tiny*) was by far the most frequent with a raw frequency of 5127 and a normalized frequency of 51.27 per million words, which constituted 97.92 percent of all the occurrences of the lemma *tiny*. This was followed by the superlative form (*tiniest*) which constituted 1.91 percent with a raw frequency of 100 and a normalized frequency of 1 per million words. Only 9 instances of the comparative form (*tinier*) were found in the BNC.

**Table 2.** *Frequencies of the Different Forms of the Lemma Tiny*

	No. of Tokens	Normalized Count (per million)	Percentage
Tiny	5127	51.27	97.92%
Tinier	9	0.09	0.17%
Tiniest	100	1	1.91%
Total	5236	52.36	100%

### Frequency of the Lemma *Small*

As with *tiny*, the lemma *small* was found to be used most frequently in its positive degree (small) with a raw frequency of 41845 and a normalized frequency of 418.45 per million words, which constituted 83.73 percent of all the occurrences of the lemma *small*. With a raw frequency of 7101 and a normalized frequency of 71.01, the comparative form (smaller) was identified as the second most frequent form of the lemma *small*, constituting 14.21 percent. The superlative form (smallest) was the least frequent with a raw count of 1032 and a normalized count of 10.32, resulting in only 2.06 percent of all the occurrences of the lemma *Small*. Table 3 presents raw frequencies, normalized counts, and percentage of various forms of the lemma *small*.

**Table 3.** *Frequencies of the Different Forms of the Lemma Small*

	No. of Tokens	Normalized Count (per million)	Percentage
Small	41845	418.45	83.73%
Smaller	7101	71.01	14.21%
Smallest	1032	10.32	2.06%
Total	49978	499.78	100%

### Frequency of the Lemma *Minute*

The lemma *minute* was found only in positive and superlative forms. No instances of the use of its comparative form were found in the BNC. It was most commonly found in its positive form (minute) with a raw count of 817 and a normalized count of 8.17 per million words. Hence, 97.03 percent of all the occurrences of the lemma *minute* consisted of the positive form. 25 instances of the superlative form (minutest) were also found in the BNC, which constituted 2.97 percent of all the occurrences of the lemma *minute*. Table 4 summarizes these findings.

**Table 4.** *Frequencies of the Different Forms of the Lemma Minute*

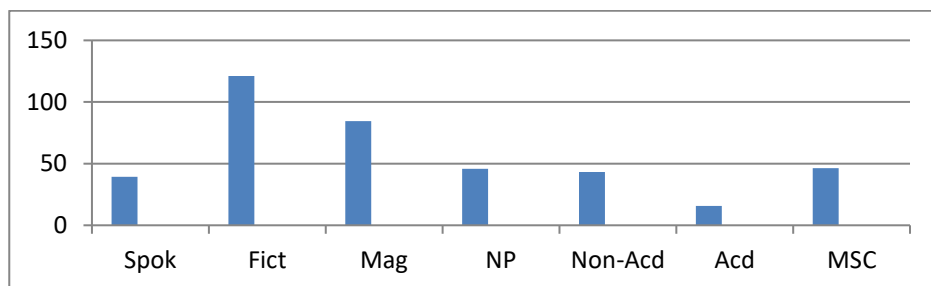
	No. of Tokens	Normalized Count (per million)	Percentage
Minute	817	8.17	97.03%
Minutest	25	0.25	2.97%
Total	842	8.42	100%

Tables 2-4 highlight many similarities and differences among the use of the different forms of lemmas *tiny*, *small*, and *minute*. One of the most notable similarities is that the positive forms of the three selected adjectives are by far the most frequent forms in the case of lemmas of all three adjectives. However, the three adjectives differ in the percentage of use of these positive forms. Positive forms of *tiny* and *minute* were almost identical with 97.92 percent and 97.03 percent, respectively, whereas the positive form of *small* was found with a percentage of 83.73. The comparative form of the lemma *small* was quite frequent (14.21%) as compared to those of the lemmas *tiny* (0.17%) and *minute* (0%). The three adjectives were similar in the use of their superlative forms with a percentage of 1.91, 2.06, and 2.97 for tiniest, smallest, and minutest, respectively.

### Register-wise Distribution of *Tiny*, *Small*, and *Minute*

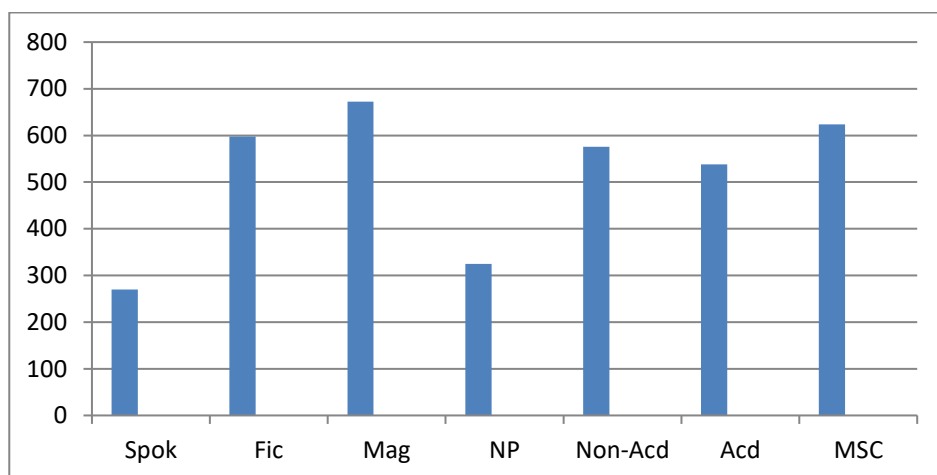
As mentioned earlier, the British National Corpus consists of seven sub-registers including Spoken, fiction, magazine, newspaper, non-academic, academic, and miscellaneous. The three selected adjectives were also compared based on their register-wise distribution in the BNC. This was done by comparing their normalized frequencies since the total number of words in each register is not the same.

**Register-wise distribution of the lemma *tiny*.** The most frequent use of the lemma *tiny* was found in fiction with a normalized frequency of 121 per million words. With a normalized frequency of 84.57 per million words, the second most frequent use of *tiny* was found in magazine. Interestingly, it occurred almost with the same frequency in spoken, newspaper, non-academic, and miscellaneous with normalized frequencies of 39.3, 45.7, 43.25, and 46.45 per million words, respectively. With a frequency of 15.8, it was found to be the least commonly used in academic writing. Figure 2 presents a visual picture of these findings.



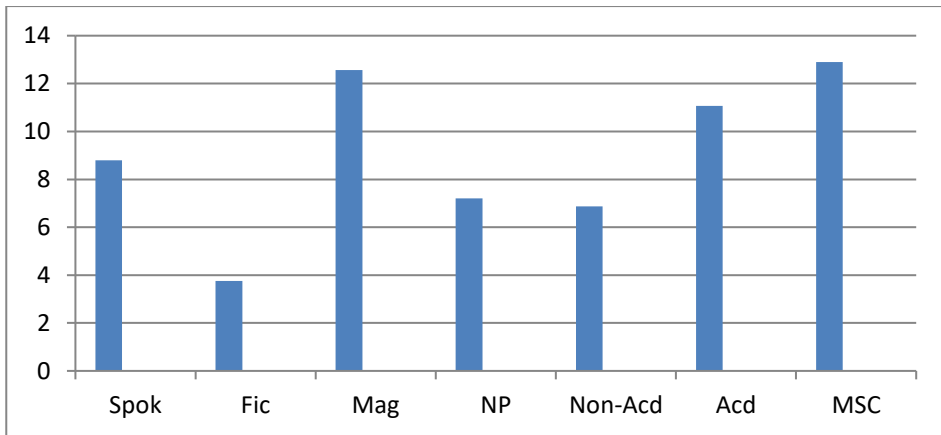
**Figure 2. Register-wise Distribution of the Lemma *Tiny***

**Register-wise distribution of the lemma *small*.** Amongst the seven sub-registers of the BNC, the lemma *small* was found to be most commonly used in magazine with a normalized count of 672.28 per million words. After this register, it was almost equally frequent in miscellaneous, fiction, and non-academic, with normalized frequencies of 623.6, 597.37, and 576 per million words, respectively. Moreover, it was fairly common in academic writing with a normalized count of 537.8 per million words. Furthermore, with a frequency of 269.9 per million words, *Small* was least frequently used in the spoken register. Figure 3 presents the summary of register-wise distribution of the lemma *small* on the basis of the normalized frequencies.



**Figure 3. Register-wise Distribution of the Lemma *Small***

**Register-wise distribution of the lemma *minute*.** The lemma *minute* was found to be most commonly used in miscellaneous and magazine with normalized counts of 12.90 and 12.57 per million words, respectively. Moreover, it was also fairly frequent in academic, spoken, newspaper, and non-academic with normalized frequencies of 11.07, 8.8, 7.20, and 6.87 per million words, respectively. With a frequency of 3.75 per million words, *minute* was observed to be used least frequently in fiction. A summary of these findings can be seen in figure 4.



**Figure 4. Register-wise Distribution of the Lemma *Minute***

A close look at figures 2-4 can help compare register-wise distribution of the lemmas of *tiny*, *small*, and *minute*. While *tiny* is most frequently used in fiction, the most frequent instances of *small* and *minute* were found in magazine and miscellaneous, respectively. Furthermore, *tiny* shows almost similar distribution in spoken, newspaper, non-academic, and miscellaneous, whereas fiction, non-academic, and miscellaneous display almost similar frequencies of *small*. Similarly, *minute* shows fairly similar distribution in newspaper and non-academic. The three adjectives also differ in their least frequent use. While *tiny* is least frequently used in academic, *small* and *minute* were found to be used least frequently in spoken and fiction, respectively. In case of comparison between spoken (an informal register) and academic writing (very formal), it can be seen that *tiny* is more than twice as frequent in spoken as it is in academic writing. Conversely, *small* is almost twice as frequent in academic as it is in spoken. This means that *tiny* is more associated with informal registers, whereas *small* tends to be more common in formal ones. Interestingly, *minute* is fairly frequent in both spoken and academic, although it is a little more common in academic than in spoken.

### **Most Frequent Collocates of *Tiny*, *Small*, and *Minute***

This section discusses the ten most frequent collocates of the three selected adjectives based on both raw frequency and MI score.

**Most frequent collocates of *tiny*.** Table 5 presents two separate lists of the ten most frequent immediate right collocates of the lemma *tiny*. The first list (on the left) displays the ten most frequent collocates of *tiny* on the basis of their raw frequencies in the BNC. The list only includes content words. Function words and punctuation marks, although

frequently collocating with *tiny*, were not taken into consideration. The second list (on the right) consists of the ten most frequent collocates of *tiny* based on their Mutual Information (MI) score.

**Table 5.** *Collocates of the Lemma Tiny*

No.	Collocate	Raw Freq.	No.	Collocate	MI Score
1	Bit	96	1	Tots	12.24
2	little	94	2	Specks	11.02
3	minority	57	3	Rowland	10.67
4	village	53	4	Speck	10.23
5	fraction	47	5	Tot	10.00
6	amount	38	6	Oblong	9.93
7	pieces	34	7	Hamlet	9.24
8	Room	34	8	Fraction	9.23
9	island	32	9	Galley	8.61
10	Part	32	10	Pores	8.37

As far as the most common right collocates of *tiny* on the basis of raw frequency are concerned, table 5 shows *bit* and *little* to be its most common right collocates. The rest of the collocates are also quite frequent, although not as much as the first two. The second list, based on MI scores, has quite a bit of rare words, such as *Rowland*, *oblong*, *galley*, etc. Surprisingly, *fraction* is the only collocate of *tiny* which is common in both frequency-based and MI-based lists, although its ranking is different in both lists. In the frequency-based list, it is ranked fifth, whereas in the MI based list, it is ranked eighth.

**Most frequent collocates of *small*.** Table 6 presents two separate lists containing the ten most frequent immediate right collocates of the lemma *small* in the BNC on the basis of raw frequency and MI score. The frequency-based list only includes content words. This list was generated after excluding function words and punctuation marks that frequently collocate with *small*. The second list consists of the ten most frequent collocates of *tiny* on the basis of their MI scores.

**Table 6.** *Collocates of the Lemma Small*

No.	Collocate	Raw Freq.	No.	Collocate	MI Score
1	number	1020	1	Imbricating	10.06
2	group	724	2	Intestine	9.57

3	amount	454	3	Mammal	9.45
4	groups	453	4	Cetaceans	9.29
5	businesses	451	5	Mercies	9.17
6	firms	451	6	Self- administered	8.97
7	proposition	402	7	Scale-like	8.65
8	scale	359	8	Crustacea	8.53
9	companies	330	9	Bowel	8.23
10	part	315	10	Tortoiseshell	8.20

In terms of raw frequency, *number* was observed to be the most common collocate of *small*. The whole list contains familiar words. The MI-based list, on the other hand, includes many unfamiliar words including *imbricating*, *cetaceans*, *crustacea*, etc. There is not a single collocate of *small* which is part of both the lists. This shows, as discussed earlier, that MI tends to rank rarely occurring collocates higher.

**Most frequent collocates of *minute*.** In the BNC, *walk* was found to be the most frequent right collocate of *minute* as an adjective on the basis of raw frequency, followed by *period*, *intervals*, *cheap*, *video*, etc. However, *equilibration* was ranked the topmost right collocate of *minute* on the basis of MI score. Surprisingly, five collocates of *minute* were common in the two lists, though with different rankings. *Walk* is ranked first in the frequency-based list, while it is fifth in the MI-based list. Interestingly, *intervals* is the third most frequent collocate of *minute* in both the lists. Furthermore, *cheap*, *amounts*, and *stroll* are ranked fourth, eighth, and tenth in the frequency-based list and sixth, ninth, and second in the MI-based list. Table 7 presents the top ten collocates of *minute* in terms of their raw frequencies and MI scores.

**Table 7.** *Collocates of the Lemma Minute*

No.	Collocate	Raw Freq.	No.	Collocate	MI Score
1	walk	67	1	equilibration	14.04
2	period	28	2	stroll	11.09
3	intervals	27	3	intervals	10.90
4	cheap	17	4	slot	9.65
5	video	13	5	walk	9.59
6	detail	13	6	cheap	9.02
7	drive	12	7	collections	8.81
8	amounts	11	8	quantities	8.75

9	flight	10	9	amounts	8.73
10	stroll	9	10	samples	8.17

A close look at the frequency-based lists of immediate right collocates of *tiny*, *small*, and *minute* reveals that there are some words which frequently collocate with at least two of the three selected adjectives. For instance, *part* is the tenth most frequent collocate of both *tiny* and *small*. Similarly, *amount* is the sixth frequent collocate of *tiny*, third frequent collocate of *small*, and the plural form *amounts* is the eighth frequent collocate of *minute*. This means that these three adjectives have some similarities in terms of meaning. Surprisingly, there is no common collocate of these adjectives as far as the MI-based lists are concerned.

### **Common Meanings and Senses Associated with *Tiny*, *Small*, and *Minute***

Stubbs (2002) is of the view that there is always a strong semantic association between a node and its collocates. Biber, Conrad, and Reppen (1998) claim that each collocate of a word has a strong tendency to be associated with a single sense or meaning, although more than one collocate may be associated with the same sense. This section discusses and compares some of the most common meanings and senses associated with the three selected adjectives by examining the ten most frequent collocates of each adjective on the basis of their raw frequencies. Lists of collocates of the three selected adjectives based on their MI scores were not taken into consideration since these lists contain many rare words.

**Common meanings associated with *tiny*.** Generally, *tiny* was found to mean the less size or quantity. However, it gives specific meanings when occurring with a certain collocate. First, In the BNC, *tiny* is frequently found as an adjective modifier, intensifying the lesser size or quantity. In this use, it most frequently collocates with *bit* and *little*. This meaning can be clearly seen in the following examples.

- 1) A tiny little thing of the size of a tomato house. <s\_conv>
- 2) Not too much milk and a tiny bit of sugar. <s\_conv>

Second, *tiny* also refers to a small group of people or fewer instances which are part of a larger population. It conveys this sense especially when it collocates with *minority*. For example,

- 3) All but a tiny minority of authorities have set affordable budgets and council taxes. <W\_newsp\_other\_report>



Third, it denotes something which is small in size or unimportant. For instance,

- 4) My mother was an innocent young country lass, born and brought up in a tiny village in the far north of Scotland.  
<w\_fict\_prose>
- 5) It was a tiny room shaped rather like a ridge tent.  
<w\_fict\_prose>

In example 4, *tiny village* can refer to a village that is small in size or an unimportant village or both. However, in example 5, it clearly refers to a small size.

Fourth, it also conveys the sense of a lesser quantity or amount. The following example illustrates this point.

- 6) Some contain only cheap vegetable oils with just a tiny amount of fragrance. <w\_pop\_lore>

*Tiny amount of fragrance* means a lesser quantity of fragrance.

**Common meanings associated with *small*.** A close look at the most frequent collocates of *small* reveals that it conveys various meanings. First, when collocated with *number*, *small* is predominantly used to refer to fewer people, institutions, groups, instances, things, etc. This meaning can be seen in examples 7 and 8.

- 7) By November, however, a small number of them had returned to work. <w\_ac\_polit\_law\_edu>
- 8) The scale of the problem was quite different in that only eight local authorities and a small number of colleges were involved.  
<w\_ac\_polit\_law\_edu>

In 7, *small* indicates fewer people, whereas in 8, it refers to fewer institutions (colleges).

Second, it is sometimes used to indicate a group of people which is not big. For example,

- 9) But the matter only came to a head when a small group of nursery nurses made a formal complaint to a course tutor.  
<w\_ac\_polit\_law\_edu>

Here, *small* refers to a group of nurses which was not large. This sense is conveyed especially when *small* collocates with *group*.

Third, *small* also refers to a quantity (both physical and abstract) which is not large. Examples 10 and 11 prove this point.

- 10) There is a small amount of water in it, but hardly any flow.  
<w\_ac\_polit\_law\_edu>
- 11) They only had a small amount of intelligence. <w\_misc>

In example 10, *small* refers to a lesser physical quantity, whereas in 11, it indicates a lesser abstract quantity.

Fourth, it is also used in financial terms. In examples 12 and 13, it is used with *businesses* and *money* in financial sense.

- 12) Small businesses are bearing the brunt of John Major's recession. <w\_newsp\_tabloid>
- 13) At the beginning of each day's trading you will be given a small amount of money (eg 20) in small change. <w\_misc>

Fifth, it also indicates something which is not large in size or unimportant or both. In 14, it refers to venues which are not of great size or are unimportant.

- 14) Productions will play at the forum for 10-14 days before touring to small scale venues in the north. <w\_newsp\_other\_arts>

Finally, *small* is used in terms of proportion of something. In example 15, *small* is used in the sense of proportion which is not large.

- 15) I have described only a small part of the whole mythological root system.

**Common meanings associated with *minute*.** A detailed analysis of the collocational meaning of adjective *minute* in the BNC revealed that, surprisingly, majority of the instances generated by the search string [minute].[j\*] were that of noun *minute*. Actually, all the examples of eight out of ten most frequent collocates of *minute* on the basis of raw frequency contain the noun. This can be seen in examples 16-20.

- 16) The centre is a thirty minute walk or a five minute taxi ride away. <w\_advert>
- 17) If there was persistent delayed transit at the end of the three minute period, the transit time was regarded as 180 seconds. <w\_ac\_medicine>
- 18) Bile was collected at 15 minute intervals.  
<w\_ac\_medicine>

- 19) Each call costs 36p a minute cheap rate.  
<w\_newsp\_tabloid>
- 20) Magistrates watched the seven minute video before  
passing sentence. <w\_news\_script>

All the above examples contain *minute* as noun, but the corpus generated these examples for *minute* as adjective. This is probably because the BNC reads noun-noun sequence as adjective-noun sequence. Hence, out of the ten most frequent collocates, the adjective *minute* was found to be collocating with only *detail* and *amounts*. The following are the two most common meanings associated with *minute*.

First, *minute* refers to a quantity which is not large as can be seen in example 21.

- 21) Even with the lids on tightly, they release minute amounts of chemicals which may be dangerous if the accumulate. <w-pop-lore>

Second, when it collocates with *details*, it refers to the finer or thorough details. For instance,

- 22) They had a passion for clothes, and every particular of how a dress or coat was made was studied in the most minute detail.  
<w-biography>

In 22, *minute* indicates finer or thorough. It is also worth noting that in this sense, *minute* is almost always used in superlative form. Table 8 summarizes the above discussed uses and meanings of *tiny*, *small*, and *minute*.

**Table 8.** *Common Meanings and Senses Associated with Tiny, Small, and Minute*

No.	<i>Tiny</i>	<i>Small</i>	<i>Minute</i>
1	As intensifier	Fewer people, institutions, groups, etc.	Finer or thorough
2	Small group of people or fewer instances which are part of a larger population	A group which is not large	
3	Not large in size or unimportant	Not large in size or unimportant	
4	A quantity or amount which is not large	A quantity (physical and abstract) which is	A quantity which is not large

---

	not large
5	In financial terms
6	In the sense of proportion

---

Table 8 presents the similarities and differences among the meanings of *tiny*, *small*, and *minute* on the basis of their most frequent collocates in the BNC. The three adjectives are similar in referring to quantities which are not large. There are more similarities between *tiny* and *small* than between either *tiny* and *minute* or *small* and *minute*. *Tiny* and *small* both refer to a size which is not small or something which is unimportant. They also refer to groups of people which do not contain a large number of people; however, *tiny* refers to the groups which are part of a larger population, whereas *small* refers to them in general. Moreover, the role of *tiny* as an intensifier is not shared by the other two adjectives. *Small*, on the other hand, gives a wide range of meanings than those given by *tiny* or *minute*. For instance, the sense of proportion and association with financial terms is peculiar to *small* only. Finally, although *minute* is associated with only two meanings, one of its meanings (i.e., finer or thorough) is not conveyed by either *tiny* or *small*. This shows that the three selected adjectives, though similar in many respects, cannot be considered strict synonyms and cannot be used interchangeably in every context.

### Comparison with Dictionary Definitions

The meanings of three selected adjectives derived from their most frequent collocates in the BNC were also compared with the definitions of these adjectives as given in *Oxford Advanced Learner's Dictionary* (2010) (henceforth OALD).

As far as *tiny* is concerned, OALD defines it as “very small in size or amount” (p. 1624). OALD mentions only two uses of *tiny* (small size and small amount), whereas our collocation-based analysis reveals that *tiny* can give four different meanings. Both the meanings of *tiny* given in OALD are captured in table 8. In addition, our collocation-based analysis indicated two more meanings (No. 1 and 2 in table 8) which are not mentioned in OALD. There are two example sentences given in OALD in which *tiny* collocates with *baby* and *minority*. Our findings reveal that *minority* is the third most frequent collocate of *tiny* and *baby* is the thirteenth most common collocate of *tiny*. This shows that OALD under-represents the range of meanings expressed by *tiny*.

In case of *small*, however, OALD lists eight uses which include a) not large in size, number, degree, amount; b) describe one size in a range of sizes of clothes, food, products used in the house; c) not as big as something else of the same kind; d) young; e) not important; f) not doing business on a very large scale; g) not written or printed as capitals; and h) not much (p. 1453). Most of these senses (senses a, c, e, f, and h) are confirmed by our analysis either directly or indirectly. However, our analysis could not identify three senses (senses b, d, and g) of *small* as given in OALD. This is because we looked at only ten most frequent collocates of *small* in our study. OALD, on the contrary, does not define *small* in the sense of proportion and referring to groups which are not large. Furthermore, Out of the thirteen collocates of *small* (*number, wedding, voice, children, boy, changes, errors, detail, achievement, farmer, businesses, cause, and hope*) given in the example sentences in OALD (p. 1453), only four occur in the list of top 100 (*number, businesses, children, and boy* ranked as first, fifth, eleventh, and nineteenth, respectively) collocates of *small* in the BNC. It means that some of the senses associated with *small* are not informed by frequency information.

For *minute*, OALD provides two meaning: a) extremely small and b) very detailed, careful and thorough which perfectly correspond to our findings. Moreover, three (*detail, amounts, and details* ranked as fifth, eighth, and fifteenth, respectively) out of five collocates of *minute* as an adjective given in OALD occur in the list of top 100 collocates of *minute* in the BNC. This may be because, as explained earlier, most of the collocates generated for *minute* as an adjective gave collocates of *minute* as a noun. Furthermore, OALD also confirms our finding that *detail* usually collocates with the superlative degree of *minute*. This shows that out of the three selected adjectives, only the meaning of *minute* given in OALD perfectly matches our findings.

## Conclusion

The present study compared the most frequent right collocates of three seemingly synonymous adjectives (*tiny, small, and minute*) in the British National Corpus. The three adjectives were compared on the basis of their frequencies of occurrence, register distribution, ten most frequent right collocates, and meanings. The meanings established with the help of the ten most frequent collocates of these adjectives were compared with their definitions given in *Oxford Advanced Learner's Dictionary* (2020).

The study shows that out of the three adjectives, *small* was found to be the most frequent in the BNC. It was also found that the positive forms of all three adjectives were most commonly used as compared to their comparative and superlative forms. The three adjectives showed great differences in terms of their register-wise distribution. As far as the meanings are concerned, although there were some similarities, each of the three adjectives conveyed certain sense(s) which were not shared by the others. Furthermore, similarities and differences were also found in the dictionary definitions of these adjectives and the meanings established on the basis of our analysis. All in all, enough differences were found among the meanings conveyed by the three selected adjectives to safely conclude that the three selected adjectives are not strict synonyms and cannot be used interchangeably in all contexts.

This study has certain limitations. First, the corpus used in this study only contains one variety of English (British English). Second, the corpus is somewhat outdated as new texts were not added to it after 1993. Third, the study compares the findings with the definitions of the selected adjectives given only in one dictionary. Fourth, it only analyzed the ten most frequent collocates of the three selected adjectives. Finally, it did not take register differences into consideration while establishing the meanings of the three selected adjectives. Future researchers should try to minimize these limitations to the possible extent.

Despite these limitations, the results of this study have important implications for lexicographers in general and for English language teachers in particular. The findings of the present study can help not only in improving the dictionary definitions of these adjectives but also help English language teachers explain these seemingly synonymous adjectives with the help of their immediate collocates and real-life examples.

## Reference

- Aldereihim, A. A. (2023). *Assessment of collocation progression in second language learners' writing at different levels of proficiency: A corpus-based study* (Doctoral dissertation, University of Southampton, UK).
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

- Coffey, S. J. (2022). English adjectives of very similar meaning used in combination: an exploratory, corpus-aided study. *Lexis. Journal in English Lexicology*, (19).
- Dang, T. N. Y., Lu, C., & Webb, S. (2022). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition*, 44(3), 708-736.
- Evert, S. (2008). Corpora and collocation. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics: An international handbook* (Vol. 2, pp. 1212–1248). Mouton de Gruyter.
- Finlayson, N., Marsden, E., & Hawkes, R. (2024). Creating and evaluating corpus-informed word lists for adolescent, beginner- to low-intermediate learners of French, German, and Spanish. *Language Teaching Research*, 13621688241288877.
- González-Díaz, V. (2021). Intensificatory tautology in the history of English: A corpus-based study. *Journal of English Linguistics*, 49(2), 182-207.
- Greaves, C., & Warren, M. (2010). What can a corpus tell us about multi-word units? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 212–226). Routledge Taylor & Francis Group.
- Johansson, M., & van Waarden, B. (2024). Structural reading: Developing the method of Structural Collocation Analysis using a case study on parliamentary reporting. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 57(3), 185-198.
- Kim, S. Y., Shin, D., & Kim, K. S. (2024). Association of lexical and collocation knowledge: A comparative analysis of a learner corpus of English and a native speaker corpus. *Vigo International Journal of Applied Linguistics*, (21).
- Kopotev, M., Kisselev, O., & Polinsky, M. (2024). Collocations and near-native competence: Lexical strategies of heritage speakers of Russian. *International Journal of Bilingualism*, 28(6), 1135-1162.

- Krishnamurthy, R. (2000). Collocation: From silly ass to lexical sets. In C. Heffer, H. Sauntson, & G. Fox (Eds.), *Words in context: A tribute to John Sinclair on his retirement*. University of Birmingham.
- McEnery, T., & Wilson, A. (2001). *Corpus linguistics: An introduction* (2nd ed.). Edinburgh University Press.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge Taylor & Francis Group.
- Meyer, C. (2004). *English corpus linguistics: An introduction*. Cambridge University Press.
- Moon, R. (2010). What can a corpus tell us about lexis? In A. O’Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 197–211). Routledge Taylor & Francis Group.
- Phillips, P., Francis, B., Webb, S., & Bull, V. (Eds.). (2010). *Oxford advanced learner’s dictionary* (8th ed.). Oxford University Press.
- Pu, P., Chang, D. Y. S., & Wang, S. (2024). Incidental learning of collocations through different multimodal input: The role of learners’ initial L2 proficiency. *System*, 125, 103416.
- Riches, N., Letts, C., Hadeel, A. W. A. D., Ramsey, R., & Dąbrowska, E. (2022). Collocational knowledge in children: a comparison of English-speaking monolingual children, and children acquiring English as an Additional Language. *Journal of Child Language*, 49(5), 1008-1023.
- Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics*, 7(2), 215–244.