A Corpus-based Study on the Meanings, Distribution, Collocations, and Formality of '*Concentrate*', '*Emphasize*', and '*Focus*'

Pariwat Imsa-ard

Faculty of Education, Ramkhamhaeng University

Abstract

This research explored the distinctions and similarities between three synonymous verbs, '*concentrate*', '*emphasize*' and '*focus*', in terms of their collocations, sense of meaning, and grammatical patterns. Data were gathered from three online dictionaries, namely (1) the Longman Dictionary of Contemporary English (2014), (2) the Merriam-Webster Dictionary (online version), and (3) the Oxford Advanced Learner's Dictionary (online version), as well as the Corpus of Contemporary American English (COCA). The study employed both quantitative and qualitative methods. The former method concerns the analysis of frequency and MI scores of each verb, while the latter method is concerned with the analysis of concordance lines. Throughout the eight text types of the COCA, '*focus*' appears to be most commonly used in academic texts and was found least often in fiction genre. Furthermore, two grammatical structures are shared by the two pairs of synonymous verbs: '*concentrate*' and '*focus*', and '*concentrate*' and '*emphasize*'. The results also revealed that, although having similar meanings, near-synonyms can behave differently in terms of collocation and semantic preferences. The results of this study can be used as supplementary materials by English language educators to help students learn more English vocabulary, especially in academic writing.

Keywords: synonymous verbs, collocation, corpus-based study, grammatical patterns

1. Introduction

The importance of vocabulary in language learning could not be overstated (Gass & Selinker, 2008). If learners cannot understand the vocabulary they are given, it is likely that they cannot comprehend it and thus fail to acquire it (Folse, 2004). To date, there is a growing body of pertinent literature which acknowledges the importance of vocabulary, especially in language learning. To understand a language more precisely, learning synonyms is thought to be an effective way to broaden one's vocabulary. However, one of the most challenging aspects in English language learning is that English language vocabulary has an infinite number of synonyms, making synonyms a challenge in vocabulary learning (Laufer, 1997). Furthermore, L2 learners even find learning synonyms confusing and difficult, since they may believe that one word is enough for them. For example, the verbs '*concentrate*', '*emphasize*' and '*focus*' are technically synonyms that correlate with the same meaning of paying attention to a particular matter. Particularly, in academic writing, these three words are frequently used.

However, when these verbs are examined in texts, they seem to have different collocational patterns which will be later discussed in the result section.

There seems to have been no analysis of the synonymous verbs '*concentrate*', '*emphasize*' and '*focus*' in a single study, and therefore, the similarities and differences of these three synonymous verbs are worth examining. Importantly, these three verbs are listed among the top 3,000 most frequent words of the BNC/COCA lists (Nation, 2017) and listed in the Academic Word List (AWL) most frequent words (Coxhead, 2000); hence, English language learners are likely to encounter these three verbs frequently. Furthermore, several researchers (e.g., Ly & Jung, 2015) argue that dictionaries' shortcomings may be one of the reasons why it is difficult for L2 learners to distinguish between loose synonyms, leading them to use synonyms in incorrect contexts. To counter these concerns, this research centered on the three English synonyms for '*concentrate*', '*emphasize*' and '*focus*', with the aim of determining some correlations and variations in their collocational patterns and senses of meaning, as well as exploring their formality. The findings of this study are intended to help teachers and educators develop vocabulary teaching materials, and help students learn how to use near-synonyms naturally.

2. Literature Review

2.1. Synonyms

A synonym is a term or phrase that has a similar or nearly identical meaning to another word. Indeed, several academics (e.g., Cruse, 2000; Jackson & Amvela, 2000; Taylor, 2003) have attempted to describe the word synonym. According to Jackson and Amvela (2000), a synonym is "a relationship of the sameness of meaning that holds between two terms" (p. 92). Furthermore, Taylor (2003) defines a synonym as a pair of terms whose semantic similarities outnumber their differences and do not primarily contrast with one another.

Synonyms are classified into two types: absolute and near synonyms. Absolute synonyms must be similar in all meanings and be interchangeable in the same way using the same collocations (Lyons, 1995; Jackson & Amvela, 2000). However, the majority of synonyms in the English language are near synonyms, also known as loose synonyms. To clarify, these synonyms have the same or identical meanings, but they cannot be used interchangeably in all cases due to a variety of reasons, including their ranging scales in meaning and use patterns.

Indeed, absolute synonyms of the same register and style are difficult to come by. Cruse (2000) shows that, in English, there might not be any true absolute synonyms. When comparing such two English adjectives as '*big*' and '*large*', they seem to have the same sense and register. According to the Longman Dictionary of Contemporary English (2014), however, '*large*' refers to something of great size and amount, while '*big*' can also refer to something of certain degree or important. When examining these two terms more closely from the comprehensive definitions in dictionaries, it can be seen that the small difference is handled by classification, with '*big*' being used for certain degrees and '*large*' being used for some amounts. As a result of this distinction, the two adjectives are not considered absolute synonyms.

2.2. How to Distinguish Synonyms

Synonyms can be distinguished using a number of criteria. The researcher differentiated synonyms based on three criteria: meaning, degree of formality, and collocation, as these are two main concerns explored in this analysis.

2.2.1. Formality

Although synonyms may have similar meanings, they may vary in terms of formality. Some words, for example, are expected to be used in formal contexts such as academic texts, while others are expected to be used in informal contexts such as colloquial language. According to Jackson and Amvela (2000), informal terms include English slang words, which are commonly used in spoken English. The following synonyms, for instance, can be differentiated by their degree of formality as follows:

Formal Language	**Informal Language**
fundamental	bare-bones
inflate	blow up
simple	a piece of cake

According to relevant literature, Bailey (2015) highlights the following characteristics of words in formal contexts:

1. The use of Latin and Greek terms such as *attitude* and *harmony*;
2. Formal language, which avoids using personal pronouns to refer to the action's performer, such as *I* and *we*;
3. The use of concise language, such as *fifteen years* rather than *several years*.
4. Accurate vocabulary, such as the distinction between the terms *law* and *regulation*.

In addition, Leech and Svartvik (2003) list the following characteristics of words in informal contexts:

1. The use of terms such as *a lot of* and *stuff* that imply imprecision;
2. The use of idioms or slang terms, such as *kids*;
3. The use of two-word verbs, such as *go ahead* and *carry about*;
4. The use of contractions, such as *can't* rather than *cannot*;
5. The use of personal pronouns like *I* believe or *we* believe;
6. Extensive use of *wh-question* types, such as *Why did the pandemic happen?* instead of *The pandemic was caused by five major factors.*

2.2.2. Collocations

In addition to formality, a collocation can be used as another criterion to distinguish synonyms. To clarify, certain synonyms are distinct due to their collocational restrictions. In other words, they are intended to co-occur in association with certain words. Collocations are described in a variety of ways by different scholars. According to Benson (1989), "Collocations are set, random repeated combinations of terms in which each word effectively maintains its meaning," (p. 85). Furthermore, Nation (2001) and O'Dell and McCarthy (2011) describe collocations as the normal form in which words occur together or belong to one another. To clarify, the adjective '*interested*' is frequently used with the preposition *in*, and the verb '*communicate*' is frequently used with the preposition *with*. According to Schmitt (2010), collocations can be divided into two types: *grammatical collocations* and *lexical collocations*.

To begin, *grammatical collocations* consist of content and grammatical words (specifically prepositions) as follows:

- Noun + preposition: *decision on/about*
- Verb + preposition: *listen to*
- Adjective + preposition: *interested in*

Second, lexical collocations are made up of two content words as follows:

- Verb + adverb: *pay well*
- Verb + noun: *make payment*
- Adjective + noun: *prompt payment*

Nation (2001) demonstrates that collocations are important in language learning for some reasons. To begin with, stored word sequences are fundamental language knowledge and use, so language knowledge can be considered 'collocational knowledge'. Second, knowing collocations is crucial to becoming fluent in English and using acceptable and appropriate language. Also, according to Kozlowski and Seymour (2003), collocation learning has some implications; for example, collocations enable EFL/ESL learners to effectively use words in conveying accurate meaning. Furthermore, appropriate collocations assist EFL/ESL students in improving their writing by allowing them to use more natural language.

2.2.3. Grammatical Patterns

To distinguish synonyms, Phoocharoensil (2010) suggests that, despite having the same central meaning, words in a series of synonyms do not all have the same grammatical patterns. According to the Oxford Advanced Learner's Dictionary (online version), the following two phrases, '*consist of*' and '*be made up of*', denote the same literal meaning, which is "to be formed from the people or things mentioned". These two words, on the other hand, cannot be found and used in the same grammatical patterns. It is correct to use '*consist of*' in active constructions, such as "This book consists of five main parts," but it is ungrammatical to use it in the passive voice, as in "This book is consisted of five main parts." Similarly, the phrase '*made up of*' is only used in passive structures, as in "This book is made up of five main parts," and it is never used in active speech, as in "This book makes up of five main parts."

2.3. Previous Studies Related to Synonyms

Several studies have looked at synonyms using corpus-based data from a variety of corpora, such as the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). According to previous research, common criteria used to distinguish synonyms include definitions and their senses of meanings, collocations, and formality.

Bergdahl (2009) looked at the distinctions between three English synonyms: '*beautiful*', '*handsome*', and '*good-looking*'. COCA and five contemporary dictionaries were used to compile the data. Despite the fact that the three synonymous adjectives communicate the same core sense of meaning, the results showed that they are used differently. For example, the adjective '*beautiful*' is more commonly used by women than by men, while the adjective '*handsome*' is more commonly used by men. According to COCA, despite the fact that the word '*good-looking*' is gender-neutral, it is more widely used by men.

Gu (2017) conducted a corpus-based analysis to compare the two English synonyms '*obtain*' and '*gain*' in terms of genre, colligation, collocation, and semantic prosody. The information was gathered from three online corpora: Sketch Engine, BYU-BNC, and Just the World. It was discovered that '*obtain*' appeared frequently in written English in the fields of natural and pure science, while '*gain*' appears frequently in trade and finance. In terms of colligation, '*obtain*' and '*gain*' share a common grammatical pattern of *verb+object* noun; however, '*obtain*' was commonly formed in passive construction. It was found that '*obtain*' was frequently formed in passive construction. Furthermore, '*obtain*' was often used to convey a concrete meaning, while '*gain*' was often conveyed metaphorically.

In addition, Petcharat and Phoocharoensil (2017) investigated the definitions, collocations, degrees of formality, and grammatical patterns of the three English synonymous adjectives '*appropriate*', '*proper*', and '*suitable*'. Three dictionaries and COCA were used to compile the data. While having the same core meaning, the three adjectives varied in collocations, degrees of formality, and grammatical patterns, according to this analysis. To clarify, it appears that '*appropriate*' is more widely used in formal settings. In terms of grammatical patterns, the words '*appropriate*' and '*suitable*' may appear in all grammatical structures of the word '*proper*', while the word '*proper*' cannot appear in all grammatical structures of the words '*appropriate*' and '*suitable*'. For instance, the word '*appropriate*' may appear in all grammatical structures of the word '*proper*' as in 'linking v. + *appropriate/proper*+ that', and the words '*suitable*' may appear in all grammatical structures of the word '*proper*' as in 'linking v. + *proper/suitable*+ that'.

Lately, Nugroho (2018) attempted to examine two presumably synonymous words, namely '*rich*' and '*wealthy*', using BNC. According to this study, '*rich*' and '*wealthy*' are considered near-synonyms because they have a similar meaning but are not interchangeable in some contexts. Furthermore, '*rich*' is a term that is widely used to describe both human and non-human beings.

The aforementioned studies demonstrate how to distinguish synonyms using various criteria and corpora and dictionaries as sources. The aim of this study was to examine the differences and similarities between the synonymous verbs '*concentrate*', '*emphasize*', and '*focus*', focusing on distribution across eight genres in COCA, and the analysis of collocations based on frequencies and mutual knowledge (MI) scores. Therefore, the aim of this study was to compare and contrast three English synonyms based on the following questions:

1. How are the synonymous verbs '*concentrate*', '*emphasize*', and '*focus*' distributed across genres?
2. What are the common noun and adverb collocates of the synonymous verbs '*concentrate*', '*emphasize*', and '*focus*'?
3. What are some grammatical patterns shared by the three synonymous verbs '*concentrate*', '*emphasize*', and '*focus*'?

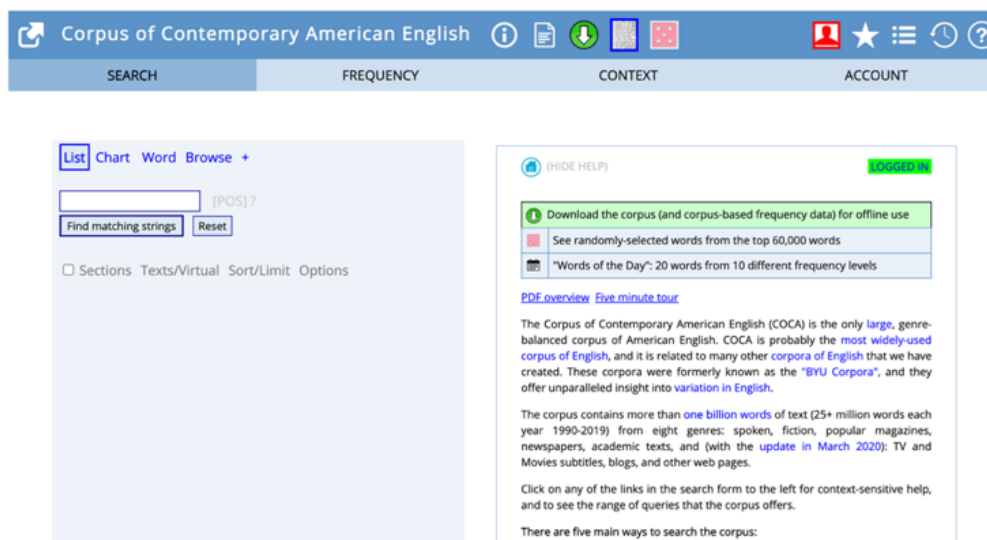## 3. Methodology

3.1. Data Collection and Data Analysis

Figure 1. Corpus of Contemporary American English

COCA, which is a corpus of American English, was used to gather data for this study (see Figure 1). COCA was selected because it is likely one of the most commonly used English corpora in the field of English language teaching, with over 1 billion words of text included from over the last two decades (Davies, 2010). Currently, COCA includes spoken, TV and movie subtitles, blogs, webpages, fiction, mainstream magazines, newspapers, and academic texts (see Figure 2). For a variety of purposes, COCA is thought to be a valuable and resourceful corpus (Davies, 2010; Hu, 2015). To begin with, COCA represents American English and is significantly larger than other available American English corpora (e.g., Corpus of Historical American English [COHA], Corpus of American Soap Operas, and TIME Magazine Corpus) due to its immense size of 1 billion words. Second, the COCA is made up of information gathered from texts that are evenly divided into eight genres. This year-to-year genre balance corpus will provide data that is distinct from other corpora. Third, COCA is a 'monitor' corpus, which means that it is constantly updated, and new texts are added.

Figure 2. Word distribution across eight genres

To answer the first two research questions, COCA was used to look up the frequencies and distribution of the three synonymous verbs '*concentrate*', '*emphasize*', and '*focus*' across eight different genres. Noun collocates and adverb collocates often co-occurring with the three synonymous verbs were investigated for the second research question. The Mutual Information (MI) value was used to measure the collocational strengths in order to select common collocates. The MI value governs "the chance of two words co-occurring in view of their frequencies in co-occurring with all the others in the corpus" (Lui & Lei, 2018, p. 6; Phoocharoensil, 2020). It is important to acknowledge, however, that the MI value has some limitations. While some collocations have high MI ratings, according to Cheng (2012) and Szudarski (2018), they may not be representative due to their low number of occurrences. Thus, both frequencies and MI values were used to assess the strength of collocation in this study. The noun collocates in COCA's top-20 frequency list with a MI score ≥ 3 (see Figure 3) were chosen because this is a significant value for collocational association (Cheng, 2012; Gablasova et al., 2017). To answer the third research question which concerns with grammatical patterns, a qualitative method was conducted. To elaborate, the grammatical structures of the three verbs were investigated in this study through analyzing the dictionaries and the concordance lines from COCA.

Figure 3. Collocations with their MI scores

Concerning the formality, the degree of formality of each concordance line in each word was examined from the Key Word in Context (KWIC) which is considered the common format for concordance lines as seen in Figure 4. As shown in Figure 4, the left-hand column lists the various genres from which the data in each line were extracted. It is worth noting that the node provide is centered on the line, surrounded by color-coded words marked according to their parts of speech; for example, pink represents verbs, bright blue represents nouns, and green represents adjectives. Concordances using color-coded KWIC allow researchers to more readily analyze some grammatical patterns (O'Keefe et al., 2007).

Figure 4. Example of the KWIC concordance line of '*concentrate*'

4. Results and Discussion

In order to answer the three research questions, the similarities and differences in meanings are demonstrated first, then the results concerning the overall frequency of the three synonymous verbs in the eight genres, the degree of formality in different contexts, and their collocations are presented, respectively.

4.1 Similarities and Differences in Meaning between 'concentrate', 'emphasize', and 'focus'

When comparing the definitions of '*concentrate*', '*emphasize*', and '*focus*' from the Longman Dictionary of Contemporary English (2014), the Merriam-Webster Dictionary (online version), and the Oxford Advanced Learner's Dictionary (online version), the three verbs are described as paying attention to a particular matter.

Table 1. Definition of *concentrate, emphasize*, and *focus* from three dictionaries

Words	Longman Dictionary of Contemporary English	Merriam-Webster Dictionary	Oxford Advanced Learner's Dictionary
concentrate	*to think very carefully about something that you are doing (V.i)*	*to bring or direct toward a common center or objective (V.t)*	*To give all your attention to something and not think about anything else*
	For example, 1) I tried to read a few pages, but I found it hard to **concentrate**. 2) The study **concentrated** on physics, not biology,	For example, 1) The lenses **concentrate** sunlight.	For example, 1) I can't **concentrate** with all that noise going on.
	to make a substance or liquid stronger by removing some of the water from it (V.t)	*to focus one's powers, efforts, or attention (V.i)*	*to bring something together in one place (V.t)*
	For example, 1) Plant roots **concentrate** uranium in cell walls.	For example, 1) Farmers are **concentrating** on the wheat harvest.	For example, 1) We need to **concentrate** resources on the most run-down areas.
		to make less dilute (V.t)	*to increase the strength of a substance by reducing its volume, for example by boiling it*
		For example 1) You should **concentrate** syrup.	For example, 1) Plants and microorganisms can **concentrate** metals from the environment.
emphasize	*to say something in a strong way*	*to give special attention or importance to*	*to give special importance to something*
	For example, 1) The report **emphasizes** the importance of improving safety standards.	For example, 1) He **emphasized** the word "maybe."	For example, 1) His speech **emphasized** the importance of attracting industry to the town
focus	*to give special attention to one particular person or thing, or to make people do this*	*to cause to be concentrated (V.t)*	*o give attention, effort, etc. to one particular subject, situation or person rather than another*

For example, 1) He needs to **focus** more on his career.	For example, 1) They **focused** their attention on the most urgent problems	For example, 1) She was distracted and finding it hard to **focus**.
to point a camera or telescope at something, and change the controls slightly so that you can see that thing clearly	*to concentrate attention or effort (V.i)*	*to change something so that you can see things clearly*
For example, 1) She turned the camera and **focused** on Martin's face.	For example, 1) **focus** on the most pressing needs	For example, 1) It took a few moments for her eyes to **focus** in the dark.

Table 1 shows that '*concentrate*', '*emphasize*', and '*focus*' tend to have a similar central sense of paying attention to a particular matter, although there are some distinctions concerning exact meanings and uses. The words '*focus*' and '*concentrate*', for example, show more precision in terms of their definitions. To illustrate, '*concentrate*' can refer to making something less dilute, and '*focus*' can refer to adapting something to see things clearly. In terms of use, '*emphasize*' are used with the object (transitive verb) whereas '*concentrate*' and '*focus*' can be used either *with* or *without* the object (transitive and intransitive verbs), depending on the meaning.

Clearly, the results of this study showed that the terms '*concentrate*', '*emphasize*', and '*focus*' do not have the exact precise definitions and thus cannot be used interchangeably in all situations. In terms of meanings, this observation is consistent with Cruse (2000) and Jackson and Amvela (2000), who indicate that two synonyms can be loose or near synonyms.

4.2 Frequency and Distribution

To gain a better understanding of general use patterns of the three synonymous verbs, the researcher initially queried COCA for overall and cross-register frequency data for each of the synonymous verbs as shown in Table 2 and Table 3, followed by their collocates in Table 4 and Table 5.

Table 2. Overall frequency of the three verbs across the genres from COCA

Genre	*concentrate*		*Emphasize*		*focus*		**Total frequency**
	Frequency	**Per million**	**Frequency**	**Per million**	**Frequency**	**Per million**	
Blog	1,249	9.71	1,123	8.73	17,892	139.11	
Web	1,282	10.32	1,317	10.60	16,292	131.12	
TV	1,944	15.18	137	1.07	5,834	45.55	
Spoken	981	7.78	1,238	9.81	14,627	115.96	
Fiction	1,956	16.53	382	3.23	5,259	44.45	
Magazines	2,234	17.72	1,476	11.71	15,124	119.95	
News	1,328	10.91	1,177	9.67	13,408	110.13	
Academic	1,808	15.09	4,656	38.87	30,721	256.46	
Total	**12,782**		**11,506**		**119,157**		**143,445**

As seen in Table 2, the verb '*focus*' occurs with the greatest frequency (119,157 tokens), whereas the verbs '*concentrate*' and '*emphasize*' are less frequent (12,782 and 11,506 tokens, respectively) in the corpus data.

Table 3. Distribution of the three verbs across the genres

Genre	*concentrate* Frequency	Per million	Genre	*emphasize* Frequency	Per million	Genre	*focus* Frequency	Per million
Magazines	2,234	17.72	Academic	4,656	38.87	Academic	30,721	256.46
Fiction	1,956	16.53	Magazines	1,476	11.71	Blog	17,892	139.11
TV	1,944	15.18	Web	1,317	10.60	Web	16,292	131.12
Academic	1,808	15.09	Spoken	1,238	9.81	Magazines	15,124	119.95
News	1,328	10.91	News	1,177	9.67	Spoken	14,627	115.96
Web	1,282	10.32	Blog	1,123	8.73	News	13,408	110.13
Blog	1,249	9.71	Fiction	382	3.23	TV	5,834	45.55
Spoken	981	7.78	TV	137	1.07	Fiction	5,259	44.45
Total	**12,782**			**11,506**			**119,157**	

In Table 3, both '*focus*' and '*emphasize*' have the highest degree of formality, as their occurrences are the highest and most frequent (30,721 tokens with 256.46 per million and 4,656 tokens with 38.87 per million, respectively) in academic texts. Regarding formality, both '*focus*' and '*emphasize*' seem to be formal, as indicated by their frequency in the academic genre. However, concerning the tokens with numbers per million, '*focus*' appears to be more commonly used in academic genres than '*emphasize*'. The fact that '*focus*' and '*emphasize*' are mainly associated with formal English is supported by them having the lowest frequency in informal contexts. Specifically, '*focus*' occurs with the lowest frequency in fiction (5,259 tokens with 44.45 per million) and TV and movie subtitles (5,834 tokens with 45.55 per million), and '*emphasize*' occurs with the lowest frequency in TV and movie subtitles (137 tokens with 1.07 per million) and fictions (382 tokens with 3.23 per million); all of which are considered to be informal English genres (Moon, 2010). Concerning the verb '*concentrate*', its occurrences are the highest and most frequent (2,234 tokens with 17.72 per million) in somewhere between formal and informal texts, that is, magazine genre (Davies, 2020). In addition, the second and third most frequent genres in which '*concentrate*' occurs are Fiction (1,956 tokens with 16.53 per million) and TV (1,944 tokens with 15.18 per million), which are considered as informal genres.

Clearly, the findings are in line with pertinent literature suggesting the adoption of genre differences as a criterion to distinguish synonyms (Jackson & Amvela, 2000; Phoocharoensil, 2020). To elaborate, the three synonymous verbs '*concentrate*', '*emphasize*', and '*focus*' commonly used in different genres, signaling that they are near-synonyms and cannot be used interchangeably in every context.

Table 4. Nouns collocates of *concentrate*, *emphasize*, and *focus* in COCA

Rank	*concentrate* Noun collocate	Frequency	MI Score	*Emphasize* Noun collocate	Frequency	MI Score	*focus* Noun collocate	Frequency	MI Score
1	effort	463	3.38	importance	1,613	6.81	attention	4,797	4.56
2	wealth	145	4.05	need	886	4.13	aspect	713	3.11
3	difficulty	130	3.62	role	521	3.33	prevention	261	3.13
4	breathing	77	4.48	approach	319	3.30	lens	244	3.16
5	inability	72	4.72	aspect	315	4.14	america	223	6.20

6	america	30	5.72	skill	298	3.21	laser	207	3.38
7	europe	30	8.28	learning	181	3.28	washington	206	8.06
8	china	23	3.63	theme	153	3.44	beam	165	3.07
9	iraq	22	9.30	curriculum	141	4.03	iraq	161	9.76
10	california	20	6.98	teaching	117	3.08	china	106	3.42
11	firepower	17	5.10	prevention	84	3.69	california	95	6.81
12	washington	17	6.87	diversity	72	3.16	europe	94	7.51
13	africa	17	7.39	significance	69	3.49	africa	93	7.42
14	schoolwork	16	5.79	cooperation	60	3.19	russia	68	7.84
15	japan	16	7.81	literacy	52	3.48	fundamental	59	3.88
16	russia	16	8.17	similarity	52	3.93	chicago	55	6.47
17	Texas	15	7.10	continuity	51	4.74	texas	49	6.39
18	Clinton	14	7.65	necessity	50	3.48	japan	49	7.01
19	germany	13	7.13	unity	50	3.52	mexico	48	6.37
20	England	11	6.82	orientation	44	3.05	november	47	7.05

According to Table 4, it is evident that '*concentrate*' and '*focus*' seem to be more synonymous, as these two words share certain common collocates. The 10 noun collocates that '*concentrate*' and '*focus*' share are America, Europe, China, Iraq, California, Washington, Africa, Japan, Russia, and Texas, where all the 10 noun collocates are related to geographic vocabulary. However, the only two noun collocates that '*emphasize*' and '*focus*' share are *aspect* and *prevention*, while '*concentrate*' and '*emphasize*' do not share any noun collocates from the top-20 frequency list.

Table 5. Adverbs collocates of *concentrate*, *emphasize*, and *focus* in COCA

Rank	*concentrate*			*emphasize*			*focus*		
	Adverb collocate	**Frequen-cy**	**MI Score**	**Adverb collocate**	**Frequen-cy**	**MI Score**	**Adverb collocate**	**Frequen-cy**	**MI Score**
1	instead	335	3.74	strongly	165	4.20	instead	1,771	3.72
2	highly	193	3.49	repeatedly	129	4.38	primarily	1,317	4.96
3	heavily	131	4.19	rightly	33	4.23	exclusively	871	5.69
4	mainly	97	3.99	traditionally	32	3.04	solely	784	5.40
5	solely	95	4.77	sufficiently	28	3.56	specifically	719	3.47
6	exclusively	92	4.86	continually	24	3.17	mainly	584	4.16
7	primarily	86	3.43	strategically	8	3.08	narrowly	452	6.29
8	intensely	23	4.21	selectively	7	3.53	heavily	333	3.12
9	geographically	23	5.30	persistently	3	3.09	intently	171	5.50
10	wonderfully	19	4.15	characteristically	3	3.28	intensely	156	4.55
11	intently	19	4.75				tightly	142	3.24
12	fiercely	15	3.88				sharply	121	3.01
13	densely	8	3.98				cognitively	60	5.76
14	spatially	8	5.14				singularly	49	5.24
15	understandably	7	3.19				principally	38	3.60

According to Table 5, it is evident that '*concentrate*' and '*focus*' seem to be more synonymous which is similar to those in Table 4, as these two words share certain common adverb collocates. The eight adverb collocates that '*concentrate*' and '*focus*' share are instead, heavily, mainly, solely, exclusively, primarily, intensely, and intently, where all the eight adverb collocates are related to vocabulary of degrees. However, '*emphasize*' does not appear to share any adverb collocates with '*focus*' and '*concentrate*' from the top-20 frequency list. In fact, there could be more adverb collocates shared by these three synonyms, but due to the exclusion of those with an MI score lower than 3, some may be absent from Table 5. Therefore, it is also significant to acknowledge that the shared noun collocates are limited to only those presented in Table 4 and Table 5.

The semantic preference of synonymous words may be used to classify noun collocates into categories based on their semantic similarity. According to Hunston (2002), semantic preference is described as the regular co-occurrence of a lexical item with items expressing a specific evaluative value. To put it another way, semantic preference is a connection between a group of semantically similar words and word types (Begagić, 2013). Ly and Jung (2015) demonstrated that knowledge about collocations can help decide the association ranges of terms searched, as well as the semantic relationships among their collocates.

Table 6. Semantic preference of noun collocates of *concentrate*

1.	**GEOGRAPHY**	America, Europe, China, Iraq, California, Washington, Africa, Japan, Russia, Texas, Germany, England
2.	**EXERTION**	Effort, difficulty
3.	**MISCELLANEOUS**	Wealth, breathing, firepower, schoolwork, Clinton, inability

All the noun collocates were grouped based on their semantic preference. In Table 6, the noun collocates of '*concentrate*' are categorized into three main themes, namely GEOGRAPHY, EXERTION, and MISCELLANEOUS. The majority of its noun collocates belong to GEOGRAPHY. Looking at Table 6, it is clear that the noun collocates appear to be associated with cities and countries, as illustrated below:

- It had been the victim of a sophisticated attack that had **<u>concentrated</u>** on at least 20 other companies originated in China. (GEOGRAPHY)
- Other says the city should be **<u>concentrating</u>** its efforts on selling OTB. (EXERTION)
- The Assad government is **<u>concentrating</u>** its firepower on big cities like Damascus and Aleppo. (MICELLANEOUS)

Table 7. Semantic preference of noun collocates of *emphasize*

1.	**EDUCATION**	role, approach, aspect, skill, learning, theme, curriculum, teaching, diversity, literacy, cooperation
2.	**REQUIREMENT**	importance, need, significance, necessity
3.	**CHARACTERISTICS**	similarity, unity
4.	**MISCELLANEOUS**	prevention, continuity, orientation

All the noun collocates were grouped based on their semantic preference. In Table 7, the noun collocates of '*emphasize*' are categorized into four main themes, namely EDUCATION, REQUIREMENT, CHARACTERISTICS, and MISCELLANEOUS. The majority of its noun collocates belong to EDUCATION. Looking at Table 7, it is clear that the noun collocates appear to be associated with teaching and learning, as illustrated below:

- The Outreach center **<u>emphasizes</u>** lifelong learning, interdisciplinary, project-based… (EDUCATION)
- And it's always been known as a place that **<u>emphasizes</u>** teaching and academic excellence. (EDUCATION)

- It would then **<u>emphasize</u>** the importance of staying in the faith. (REQUIREMENT)
- South Korea experiences with political liberalization may **<u>emphasize</u>** an important similarity. (CHARACTERISTICS)
- … be incorporated in a public health approach to drug policy that **<u>emphasizes</u>** prevention education and diversion to treatment … (MISCELLANEOUS)

Table 8. Semantic preference of noun collocates of *focus*

1.	**GEOGRAPHY**	America, Washington, Iraq, California, Europe, Africa, Russia, Chicago, Texas, Japan, Mexico
2.	**SCIENCE**	Aspect, prevention, lens, laser, beam
3.	**MISCELLANEOUS**	Attention, fundamental, november

All the noun collocates were grouped based on their semantic preference. In Table 8, the noun collocates of '*focus*' are categorized into three main themes, namely GEOGRAPHY, SCIENCE, and MISCELLANEOUS. The majority of its noun collocates belong to GEOGRAPHY. Looking at Table 8, it is clear that the noun collocates appear to be associated with cities and countries, as illustrated below:

- Many of Germany's top exporters have for years been **<u>focusing</u>** on China and other more dynamic markets in Asia. (GEOGRAPHY)
- Because it comes with a macro lens, the iPhone **<u>focuses</u>** on the subjects, not the background. (SCIENCE)
- We firmly believe that it is the most important place to **<u>focus</u>** public attention and individual ingenuity. (MISCELLANEOUS)

4.3 Degree of Formality

The degree of formality of the texts was assessed using Bailey's (2015) and Leech and Svartvik's (2003) main features. The results showed that COCA emphasizes the degree of formality, as seen in Table 9.

Table 9. Percentages of concordance lines of the three verbs in formal and informal contexts

Synonyms	Total concordance lines of each synonym: 200			
	Formal		Informal	
	No. of lines	Percent	No. of lines	Percent
concentrate	107	53.50%	93	46.50%
emphasize	141	70.50%	59	29.50%
focus	159	79.50%	41	20.50%

Table 9 shows the distribution of concordance lines for each synonym in both formal and informal contexts after examining the degree of formality of each concordance line (see Figure 4) using Bailey (2015) and Leech and Svartvik (2003) guidelines for separating formal and informal synonyms. The words '*concentrate*', '*emphasize*', and '*focus*' appear in both formal and informal settings. Concerning the formal contexts, '*focus*' scores the highest (159 lines) while '*emphasize*' and '*concentrate*' score less (141 lines and 107 lines, respectively).

Furthermore, the results show that the three synonymous words studied in this research are not identical in terms of formality due to the occurrences in different genres and the analysis of concordance lines based on the criteria by Bailey's (2015) and Leech and Svartvik's (2003). In addition, 'concentrate' appears in formal and informal texts at similar percentages (53.50% and 46.50%, respectively), which accords with the findings in Table 3 indicating that 'concentrate' appears in somewhere between formal and informal texts. Overall, these findings seem to be in line with the results from the distribution of three synonymous verbs across the genres presented in Table 3.

4.4 Grammatical Patterns

The grammatical structures of three verbs were investigated in this study through analyzing from the Longman Dictionary of Contemporary English (2014), the Merriam-Webster Dictionary (online version), the Oxford Advanced Learner's Dictionary (online version), and the Corpus of Contemporary American English (COCA) as shown in Table 10.

Table 10. Shared grammatical structures of *concentrate*, *emphasize*, and *focus*

Synonyms	Shared Grammatical Structures
concentrate and *focus*	*concentrate/focus* + on + N.
concentrate and *emphasize*	*concentrate/emphasize* + that + S. + V.
emphasize and *focus*	-

Table 10 shows that these three synonyms have two grammatical structures in common. The verb 'concentrate' appears to share grammatical structures with 'focus' (concentrate/focus + on + N.) and 'emphasize' (concentrate/emphasize + that + S. + V.), whereas 'emphasize' and 'focus' do not share any grammatical structures in common. A possible explanation is that 'focus' requires being followed by the preposition *on* and can be considered an intransitive verb, while 'emphasize' does not require any prepositions and can be considered a transitive verb. Nonetheless, it is likely that certain grammatical structures are missing from the dictionaries and 300 concordance lines used in this analysis and some of the grammatical structures are not shared by both synonymous verbs. The findings of this study support what Phoocharoensil (2010) suggested, namely that words in a series of synonyms cannot all be used in the same grammatical patterns while having the same core meaning.

The results of this study indicate that corpus-based research can provide more informative knowledge than a dictionary alone, especially in terms of use in different contexts (Hunston, 2002). In this study, the results show that the synonymous verbs 'concentrate', 'emphasize', and 'focus' are called 'near-synonyms,' and this is in line with Phoocharoensil (2020) and Jarunwaraphan and Mallikamas (2020), who stated that near-synonyms may be used differently in terms of collocation and semantic preferences while having identical definitions. Other linguists (e.g., Stubbs, 2001; Taylor, 2003) also back this up, showing that true synonyms are uncommon and that all synonyms are truly near-synonyms. Additionally, McCarthy et al. (2010) established that true synonyms are very uncommon and non-existent, arguing that they are uneconomical since they result in unnecessary redundancy in a language. Other research exploring synonymy (e.g., Liu, 2010; Petcharat & Phoocharoensil, 2017; Uba, 2015) showed that synonyms are often not exactly similar in context, and thus are not totally interchangeable. The analysis of these three synonyms demonstrates that, although they have

a common core meaning, they vary in certain specific features or senses of meaning, grammatical patterns, and collocations. As a teacher, I often notice that many students misunderstand these three synonymous terms. For instance, several of them use an ungrammatical pattern such as *I want to emphasize this topic* in analogy to *I want to concentrate/concentrate on this topic.* It is possible that they base their usage on an analogy, resulting in the term being used in the pattern of another synonym (Phoocharoensil, 2010).

5. Conclusion

Regarding distributions across eight different genres of the synonyms '*concentrate*', '*emphasize*', and '*focus*', it can be concluded that these three synonymous verbs cannot substitute each other in all contexts due to their different degrees of formality and collocates. '*Focus*' is the most common in formal contexts because it has the greatest tendency to appear in academic contexts. Furthermore, some verbs share certain common collocates. America, Europe, China, Iraq, California, Washington, Africa, Japan, Russia, and Texas are the 10 noun collocates that '*concentrate*' and '*focus*' share, and all 10 noun collocates are tied to geographical vocabulary. However, the only two noun collocates shared by '*emphasize*' and '*focus*' are aspect and prevention, while '*concentrate*' and '*emphasize*' do not share any noun collocates from the frequency list. It is proposed in this study that data from corpora may provide additional information not contained in dictionaries. Furthermore, this research shows that corpora can be helpful because they act as a valuable language teaching tool for English language teachers, and it is critical that teachers teach their students to be aware of the differences between synonymous terms and how to use them properly. To support this, Shahzadi et al. (2019) also advocate the use of corpus-based analysis in language instruction as a way to explain synonym usage and help language learners expand their vocabulary. As a result, using corpus methods to illustrate the difference in genres and formality is an essential and useful criterion for effectively distinguishing synonyms (Szudarski, 2018).

It should be noted, however, that this study has some limitations. First, the findings of this investigation focused primarily on interpreting facts from corpus concordance lines, and the English language may not be accurately represented by a single corpus (Hunston, 2002). Second, only three synonyms are investigated in this study. As a result, further research should focus on other terms in the collection of synonyms, such as highlight or stress. Finally, the number of noun collocates is restricted to the top 20 frequency lists. Other noun collocates beyond the top-20 frequency list can help to create a clearer and more detailed image of the collocational patterns of the synonyms being studied. Moreover, other kinds of statistics (e.g., *t*-score) may be used in collaborating with MI scores to demonstrate the degree of confidence with which we can claim the association of collocations.

The Author

Pariwat Imsa-ard (pariwat.i@rumail.ru.ac.th) is a lecturer in ELT at Ramkhamhaeng University, Thailand. He is currently pursuing his Ph.D. in English language teaching at Thammasat University, Thailand. He holds an M.A. in TESOL with merit from the University of York, United Kingdom. His areas of interest include language assessment, teacher education, and English language teaching.

References

Bailey, S. (2015). *Academic writing: A handbook for international students* (4th ed.). Routledge.

Begagić, M. (2013). Semantic preference and semantic prosody of the collocation make sense. *Jezikoslovlje*, *14*(2-3), 403-416. https://hrcak.srce.hr/112198

Benson, M. (1989). The collocational dictionary and the advanced learner. In M. L. Tickoo (Ed.), *Learners' dictionaries: State of the art* (pp. 84-91). SEAMEO Regional Language Centre.

Bergdahl, M. (2009). *Is he beautiful, handsome or good-looking? A study of putative synonymy in three adjectives* [Master's thesis, Luleå University of Technology].

Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Cruse, D. A. (2000). *Meaning in Language: An introduction to semantics and pragmatics*. Oxford University Press.

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, *25*(4), 447-464.

Davies, M. (2020). *The COCA corpus (new version released March 2020)*.

Folse, K. S. (2004). The underestimated importance of vocabulary in the foreign language classroom. *CLEAR News*, *8*(2), 1-6. http://www.seasite.niu.edu/trans/articles/Underestimated%20Importance%20of%20Vocab.htm

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence. *Language Learning*, *67*(S1), 155-179. doi: 10.1111/lang.12225

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd ed.). Routledge.

Gu, B. J. (2017). Corpus-based study of two synonyms: Obtain and gain. *Sino-US English Teaching*, *8*, 511-522.

Hu, H. M. (2015). A semantic prosody analysis of three adjective synonymous pairs in COCA. *Journal of Language and Linguistic Studies*, *11*(2), 117-131.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.

Jackson, H., & Amvela, E. Z. (2000). *Words, meaning, and vocabulary: An introduction to modern English lexicology*. Cassell.

Jarunwaraphan, B., & Mallikamas, P. (2020). A Corpus-based study of English synonyms: Chance and opportunity. *REFLections*, *27*(2), 218-245. https://so05.tci-thaijo.org/index.php/reflections/article/view/248710

Kozlowski, A., & Seymour, L. (2003). The importance of collocation in English language teaching. *Contact*, *29*(3), 41-49.

Laufer, B. (1997). What's in a word that makes it hard or easy? Intralexical factors affecting the difficulty of vocabulary acquisition. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 140-155). Cambridge University Press.

Leech, G., & Svartvik, J. (2003). *A communicative grammar of English* (3rd ed.). Routledge.

Lei, L., & Liu, D. (2018). The academic English collocation list: A corpus-driven study. *International Journal of Corpus Linguistics*, *23*(2), 216-243.

Liu, D. (2010). Is it a chief, main, major, primary, or principal concern?: A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics*, *15*(1), 56-87.

Ly, T. H., & Jung, C. K. (2015). A corpus investigation: The similarities and differences of 'cute', 'pretty' and 'beautiful'. *3L: The Southeast Asian Journal of English Language Studies*, *21*(3), 125-140. https://ejournal.ukm.my/3l/article/view/8788

Lyons, J. (1995). *Linguistic semantics: An introduction*. Cambridge University Press.

McCarthy, M., O'Keeffe A., & Walsh, S. (2010). *Vocabulary matrix*: *Understanding, learning, teaching*. Heinle.

Merriam-Webster dictionary. (n.d.). https://www.merriam-webster.com/

Moon, R. (2010). What can a corpus tell is about lexis?. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics*. (pp. 345-358). Routledge.

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

Nation, P. (2017). *Vocabulary lists*. Retrieved April 2, 2021 from https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists.

Nugroho, D. Y. (2018). A corpus linguistics investigation of two near-synonymous words: Rich and wealthy. *English Education: Journal of English Teaching and Research*, *3*(2), 118-127.

O'Dell, F., & McCarthy, M. (2011). *English collocations in use: Advanced; how words work together for fluent and natural English; self-study and classroom use*. Cambridge University Press.

O'Keefe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.

Oxford advanced American dictionary. (n.d.). https://www.oxfordlearnersdictionaries.com/definition/american_english/

Petcharat, N., & Phoocharoensil, S. (2017). A corpus-based study of English synonyms: Appropriate, proper, and suitable. *LEARN Journal: Language Education and Acquisition Research Network*, *10*(2), 10-24. https://eric.ed.gov/?id=EJ1229630

Phoocharoensil, S. (2010). A corpus-based study of English synonyms. *International Journal of Arts and Sciences*, *3*(10), 227-245.

Phoocharoensil, S. (2020). A genre and collocational analysis of consequence, result, and outcome. *3L: Language, Linguistics, Literature*, *26*(3), 1-16. https://ejournal.ukm.my/3l/article/view/38851

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Shahzadi, A., Asghar, A., & Javed, S. (2019). Effectiveness of corpus in teaching English synonyms. *Corporum: Journal of Corpus Linguistics-CJCL*, *2*(1), 51-65.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell Publishers.

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. Routledge.

Taylor, J. R. (2003). Near synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Science*, *25*, 263-284.

Uba, S. Y. (2015). A corpus-based behavioural profile study of near-synonym: Important, essential, vital, necessary, and crucial. *International Journal of English Language and Linguistics Research*, *3*(5), 9-17.