A Corpus-Based Study of English Synonyms: Small, Little, Tiny and Petite

Raeda Ammari
Amman Arab University - Jordan, raedaammari@aau.edu.jo

Wajed Al-Ahmad
Ministry of Education - Jordan, wajdrasmiahmad@gmail.com

A Corpus-Based Study of English Synonyms: Small, Little, Tiny and Petite

Raeda Ammari[1*], Wajed Al-Ahmad[2].

[1]Amman Arab University, Amman, Jordan.
[2]Ministry of Education, Amman, Jordan.

Abstract

Synonymy is one of the complex linguistic features that learners of English face as a second language. This study investigates the similarities and differences among four English synonyms—*small, little, tiny*, and *petite*—in terms of their genres, usage comparisons, grammatical patterns, and collocations. The data is derived from both the Corpus of Contemporary American English (COCA) as a native English corpus and the Cambridge Learner Corpus (CLC) as the non-native corpus. The results from both corpora show that *small* and *little* are frequently used, while *petite* is rarely used in COCA and not used at all in CLC. Additionally, *little* is mainly used in TV and fiction genres, whereas *small, tiny*, and *petite* are used in news and magazine genres. The grammatical patterns are similar across these synonyms, except for one distinct pattern associated with little in COCA. Furthermore, the outcomes revealed that there is no comparative form for *petite* in either corpus; however, *little* has two comparative forms that differ in meaning. The researchers suggested a few recommendations based on these findings.

Keywords: Synonymous Adjectives, Corpus-Based Analysis, Grammatical Patterns, Collocational Patterns, Learner Corpus.

1. Introduction

Over the past century, there has been a dramatic increase in learning a new language, especially English, as it is considered a lingua Franca. Vocabulary is one of the essential parts of speech to communicate effectively (Gass & Selinker, 2008). Generally, students often recognize the significance of learning vocabulary as they carry around dictionaries rather than grammar books (Schmitt, 2010). This study examines the use of four synonymous adjectives, namely: small, little, tiny, and petite, to draw on similarities and differences in the used patterns by native and non-native users in two corpora: COCA and CLC.

According to Webb (2007), teaching students to study English synonyms is a complex and challenging task. Learners probably confuse words that are similar in meaning rather than words that are different in their semantic links. Learners find it confusing to memorize all the synonyms as not all are used interchangeably in a similar context.

Taylor (2002) explained that synonymy is when a single meaning can be explained by more than one lexical item. Chung (2011) indicated that linguists and lexicographers are interested in identifying the substitution phenomenon among synonymous words. Accordingly, it is relatively complicated to choose appropriate synonyms in various contexts and understand the familiar way to use them, especially for L2 learners. Thus, students find synonym learning challenging as they have difficulties differentiating synonyms in terms of usage (Morimoto, 2020).

Lee and Liu (2009) stated that many previous studies have illustrated that discriminating near-synonyms requires sufficient semantic competency, and finding the definitions in dictionaries is not sufficient. Sometimes extra background information would help students grasp the intended meaning (Su, Qi, Zhan & Weidong 2020).

Although some words carry the same meaning, they might have different grammatical structures with their fixed collocations. As a result, English language learners need to learn these differences to choose the right words and communicate in English effectively (Murphy, 2010). Synonyms are a complex linguistic phenomenon. Liu (2010) argued that the notion of synonyms depends on the degree of similarity or sameness. Thus, no synonymous pairs need to be interchangeable as both items are not identical in meaning. One of its complexities is that a pair of synonyms might nearly have the same meaning but not necessarily have the same referent (Lyons, 1995). This implies that having the same referent may not qualify items to be called synonyms.

1.1 Previous Studies on Synonymous Adjectives

As the corpus-based approach of language analysis is a reliable source for empirical data, a large and growing body of literature has examined synonym differences through corpus-based data drawn from various corpora, e.g., the Corpus of Contemporary American English (COCA). Standard criteria to differentiate synonyms are used by most researchers, such as the meanings, collocations, grammatical patterns, and formality degree. This would help language teachers and learners to find differences in the use of language. Moreover, the majority of the studies focused only on corpus-based data, while others used dictionaries as another data source. In his research,

Taylor (2003) investigated a pair of near-synonyms high and tall in terms of their noun collocates. Following MacLaury's vantage theory analysis of «co-extensive» color categories, Taylor broadens his approach to the study of seemingly synonymous adjectives from the Lancaster-Oslo/Bergen corpus (LOB). Taylor tested the hypothesis by using a questionnaire asking individuals to judge the acceptability of high and tall collocation with various nouns. The outcomes indicated that high is used more frequently in the language than tall, while high is employed to a broader range of entities. Lastly, there is an overlap between the two words, and the difference is not clear-cut.

Three synonyms were examined by Bergdahl (2009), namely: beautiful, handsome, and good-looking; the data was collected from five contemporary dictionaries and the Corpus of Contemporary American English (COCA). The results demonstrated that the three words have the same core meaning, "pleasing to look at." However, beautiful is more associated with a woman rather than a man, while handsome is used more often with a man. The dictionaries are considered good-looking gender-neutral, but it is used more often with men as seen in the corpus samples. In terms of connotation, handsome is sometimes used with a woman to show masculine features.

Following the BP approach to study five near-synonyms, Liu (2010) differentiated the internal semantic structures of five adjectives chief, main, major, primary, and principal,

- ٢ -

depending on the Corpus of Contemporary American English (COCA). He focused on their distributional patterns, which are determined mainly by the types of nouns they modify. Hence, he analyzed Mutual Information (MI) scores of nouns modified by the five adjectives. The outcomes revealed that the types of nouns the five adjectives modify differ from each other. Liu conducted a test called hierarchical configural frequency analysis. Moreover, the findings revealed that corpus-based behavioral profile analysis works well in delineating the internal structure and the subtly semantic differences among the five adjectives. Besides, the co-occurrence crucial for examining the semantics of synonymous adjectives is different from those of synonymous verbs.

In a comprehensive study, Cai (2012) examined the adjective great and its six synonyms: fabulous, excellent, fantastic, terrific, wonderful, and awesome. The data was extracted from the Corpus of Contemporary American English (COCA). The findings showed that great is used more than the other synonyms due to its comprehensive meaning. Most of these synonyms, great, terrific, fantastic, wonderful, and fabulous occurred in the spoken genre, while awesome and excellent are generally used in magazines. Furthermore, synonymous adjectives have a variety of adverb collocates and few similarities among noun collocates.

Another study performed by Chanchotphattha (2013) elucidated three adjectives' linguistic information profile: different, diverse, and various. He compared them according to the formality, meaning, collocations, dialect, grammatical patterns, and dictionary definitions. The data for investigation was collected from 300 concordance lines drawn from BNC and the eighth edition of Oxford Advanced Learner's Dictionary 2010. The data analysis showed that the three adjectives are not entirely similar and cannot be interchangeable in all contexts. Their uses depended on "dialect, formality, collocations and grammatical patterns" (ibid, p.30). It was noticed that there were no significant differences regarding the dialect between diverse and various.

Furthermore, adverbs, linking verbs, prepositions, and nouns are the main words that usually collocate with these adjectives, except for a collocation comprising "various + preposition" was not found at all. (Hoffmann, 2014) performed a corpus-based study to examine six near-synonyms viz., nice, kind, lovely, friendly, gorgeous, and pleasant. The data was gathered from the definitions taken from three online dictionaries: The Collins American Dictionary, The Macmillan Dictionary, and The Merriam Webster Dictionary. Later, the similarities and differences among the dictionaries were analyzed and summarized. Depending on the COCA's frequencies, the most frequent collocates of the adjectives were categorized into lexical patterns and the stylistic variation across the different genres. The results unveiled differences in meaning among the corpus analysis and the definitions of the adjectives in these dictionaries.

In the same vein, Laurea (2014) focused on four sets of near-synonyms within different parts of speech. The three adjectives, namely obligatory, mandatory and compulsory, were examined from the syntactic and pragmatic similarities and differences. The data were collected from two resources, various monolingual, bilingual, and synonym dictionaries, as well as three corpora viz., BNC, COCA, and Collins Word banks online. The outcomes assured that although they share the same basic meaning, the three synonyms are not interchangeable in all contexts. Another distinction is the semantic features; for example, the adjective compulsory entailed punishing those who do not respect the rules while obligatory and mandatory do not. In addition, the adjective compulsory is used in specific situations to indicate that something is done to keep people safe from present danger.

- ٣ -

Nisani (2015) explored three similar adjectives in a different study: possible, probable, and likely. The data was extracted from COCA and two online dictionaries: the Oxford Advanced Learner's Dictionary and the Longman Dictionary of Contemporary English. Later, these terms were compared and contrasted according to their noun collocation, meaning, grammatical pattern, and degree of formality. Because of the differences among these adjectives in the four aspects cannot replace each other in all contexts. Moreover, language learners can enhance their vocabulary knowledge of the corpus linguistic information as it provides extra data which dictionaries don't have.

Similarly, Sirikan (2015) tackled loose synonymous adjectives through a corpus-based approach, i.e., appalling, horrific, horrible, and hideous. These synonyms were investigated to compare their grammatical patterns among the COCA (American English), BNC (British English), Oxford Advanced Learner's Dictionary, and Oxford Collocation Dictionary. The concordance lines and frequencies were taken from the corpora while other information from the dictionaries. The ten most frequent nouns with their collocate were analyzed within the first 200 concordance lines for each word. The results showed that dictionaries are insufficient to provide us with complete information. Additionally, some grammatical patterns appear less in Oxford Advanced Learner's Dictionary than in the two corpora. Besides, some of the noun collocates listed in the Oxford Collocation Dictionary do not appear at the top ten list drawn from COCA. According to the two corpora, these words cannot be used interchangeably in all contexts. For this reason, any corpus could be a valuable and beneficial complementary tool for extra-linguistic information that is not available in dictionaries.

Another interesting study was performed by Liu (2010) to identify the similarities and differences in three adjectives deadly, lethal, and fatal. The researcher examined these adjectives from the grammatical patterns, formality, and collocational profile aspects. The BNC and Oxford Online Advanced Learner's Dictionary 2010 were used to obtain the data. One hundred concordance lines for each adjective were selected and analyzed. Regarding the first parameter, the results highlighted that the most frequent grammatical pattern was (nouns + adjective) used for each synonym. Then, this pattern was followed by a "linking verb + adjective." The BNC presented extra information on grammatical patterns than the dictionary did. As for the second parameter, the analysis pointed out that all three adjectives are used in similar contexts despite their formality.

The obtained results assured that the three adjectives collocate mainly with linking verbs, nouns, adverbs, and prepositions. Nonetheless, they cannot substitute for each other in all contexts. Conversely, the two adjectives lethal and fatal are used in formal contexts more than the adjective deadly.

In the same vein, Uba (2015) concentrated in this study, the internal semantic structure of a wide set of near-synonym adjectives, namely, vital, essential, necessary, important, and crucial. The data sample was extracted from the BNC and nine traditional reference materials. The results highlighted the distributional patterns of the nouns that each adjective modifies by adopting the behavioral profile. The data analysis surprised the existing definitions of this set of near-synonym adjectives by dictionaries, which are regarded as inaccurate for some definitions.

Putri, Rajeg, and Wandia (2017) focused on the adjective admirable and its near synonyms: commendable, laudable, estimable and admirable in COCA. The researchers examined the distribution, collocational structure meanings of these four adjectives in COCA. The results indicated how each adjective has a different frequency in each genre. The highest frequency was

- ٤ -

for admirable, i.e., 500 among other adjectives, while estimable scored the lowest. The adjectives are mainly collocated with nouns as in the pattern "admirable + noun." They also have the pattern of "adverb + adjective." Their functions appeared clearly in the second type of lexical collocation, which was attributive and predictive ones.

2. Statement of the problem

Synonymy is an essential yet complicated linguistic feature because they are not interchangeable. The subtle differences in meaning change their connotations, implications, and register (DiMarco et al., 1993). Like any language in the world, English has plenty of synonyms due to borrowing and historical reasons. For this reason, English speakers can express themselves and convey various meanings more accurately and efficiently for the intended addressees and context (Liu & Espino, 2012). Unfortunately, EFL (English as Foreign Language) students consider it a thorny area because of their degree of variations in meaning and usage. Synonyms usually do not have identical meanings and are hence not completely interchangeable (Liu, 2010). Moreover, (Divjak &Gries, 2006) perceived those synonyms 'are neither free variation nor complementary distribution.' Synonyms are vital for expressing our thoughts, ideas, and feelings for effective communication (Edmonds &Hirst, 2002). Despite its importance, little attention has been given to the concept (Divjak &Gries, 2006; Edmonds &Hirst, 2002; Liu, 2010; Taylor, 2003).

2.1 Research Questions

1. What are the similarities and differences among the four synonymous adjectives, namely small, little, petite, and tiny, in terms of frequency in COCA and CLC?

2. What are the similarities and differences among the four synonymous adjectives, namely small, little, petite, and tiny, in terms of genres, comparison, and grammatical patterns in COCA and CLC?

3. What are the similarities and differences among the four synonymous adjectives, namely small, little, petite, tiny, in terms of comparison, collocations, and preceding verb in COCA and CLC?

3. Methodology

3.1 Corpora and data collection

The present study investigated two corpora: the corpus of contemporary American English (COCA) and the Open Cambridge Learner English corpus (uncoded) (CLC). The scope of the study is limited to four synonymous adjectives, namely small, little, petite, and tiny. These four adjectives were chosen from the topmost frequent words used in spoken or written English in COCA. Sinclair defined corpus or in plural' corpora" as follows: "A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research." (Sinclair, 2005:12). Hence, a corpus stands for an extensive collection of natural texts, both written and spoken, and they can be found stored in digital form. Usually, a corpus consists of millions of words from various genres: fiction and non-fiction writing, academic papers, newspaper articles, and others (Meyer, 2002).

Contemporary Corpus of American English (COCA): was released in 2008 online where it contained more than 450 million words as every year, 20 million new words were added from the period 1990-2009. Now, it includes about one billion words. COCA is the most widely used

- ٥ -

corpus from English-Corpora.org. It is a vast, diverse, and publicly available corpus of American English that contains various texts from a wide range of genres and texts. Consequently, learners, teachers, and researchers have a dynamic source for observing continuing changes in the language. (Davies, 2009).

The Open Cambridge Learner Corpus – uncoded (CLC): is a 2.975.701 million word balanced written corpus compiled by Cambridge University Press and Cambridge English language assessment and collected from Cambridge English Language Assessment exams (FCE, CAE, CPE) for more than 10,000 students. The students come from over 60 countries of 7 different mother tongues. You can access CLC through Cambridge Sketch Engine, web-based software that allows researchers to examine the Cambridge Corpus resources.

3.2. Corpus tools

This section briefly describes different corpus tools used to analyze data collected from COCA and CLC. Interestingly, each corpus is equipped with powerful data analysis tools and visualization. COCA provides several features such as a list to find out the frequencies of use of a query in the corpus in general, including the raw and norm frequencies. The chart feature shows the genre distribution of a specific query with raw and norm frequencies, providing a timeline analysis of use. A vital function researchers can find on www. English-corpora/COCA is the collocate feature that allows searching the collocations of any word. Researchers used the feature WORD for the current study, presenting various information related to a particular query.

Word feature is a dynamic and powerful function that allows researchers to find much data related to any word. The window search provides different information: part of speech, dictionary meanings from different dictionaries (Cambridge dictionary, Merriam-Webster, and other sources such as Google), distributional frequency of the query is among the eight various genres of COCA, links to some videos that illustrate how native speakers employ the query (You Glish, play phrase and Yarn). Moreover, the translation query in other languages (Translate), synonyms, clusters, topics in which the query is more commonly used, collocates (including the most frequently used words of different parts of speech such as nouns, verbs, adjectives, and adverbs), and KWICK. At the top of the window of the WORD function, the researcher can look for more detailed information about the query related to topics, collocates, dictionary, clusters, related words, concordance lines, and many more (see chart 1).
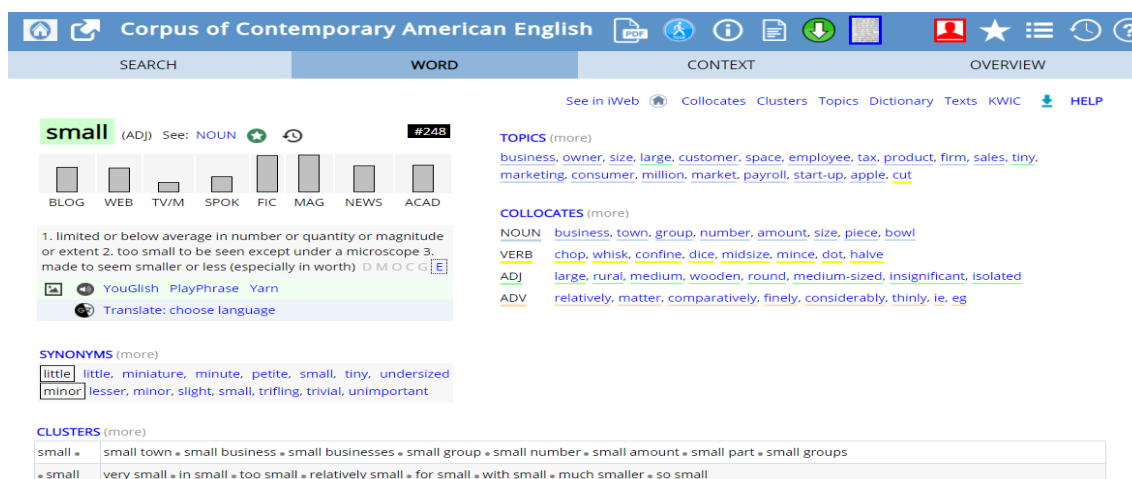

Chart (1): WORD- feature in COCA interface

- ٦ -

Cambridge Sketch Engine: Cambridge Sketch Engine provides different corpus functions such as Concordances, which is a dynamic feature that generates concordance lines of any word in the corpus, with many several ways to display information related to the target query (cf. distribution of the query in the corpus, frequency information, sorting, collocates, extended context and more). Another function is Word Sketch, which provides the researcher with information about the most frequent collocations and related words to the search query. Word Sketch Difference is a corpus tool used to compare the most used collocates with two words; such comparison is explained in two main methods: tables and visualization tools. Chart (2) shows the different corpus functions on the Cambridge Sketch Engine homepage. Each tool allows researchers to examine different linguistic behaviors related to any linguistic query (Thesaurus, N-grams, Keyword, concordance, and others). The software page also shows the search history for researchers to facilitate retrieving the data being searched.

Chart (2): Search features on Sketch Engine search page

3.3 Data Analysis

The present paper is based on quantitative and qualitative analyses of similarities and differences between the four synonymous adjectives: small, little, tiny, and petite in the Corpus of Contemporary American English (COCA) and the Open Cambridge Learner Corpus (CLC). Quantitative analysis is based on counting the raw and normalized frequencies of the queries in each corpus. Raw frequency refers to the count number of a word in a corpus, whereas normalized frequency or per million frequency refers to the number of times a word occurs in every million words in a corpus. To show comparable results, the researchers used the normalized frequency (PM). Another aspect investigated is the distributional frequencies of use of the four synonyms in the spoken and written subsections of COCA. Each subsection of COCA is calculated as the following: SPOKEN and TV/MOVIES genres represent the spoken form, whereas BLOG, WEB-GEN, FICTION, MAGAZINE, NEWSPAPER, and ACADEMIC represent the written form of COCA. This calculation is used to examine in which form of American English each adjective is more commonly used. Table (1) displays the size of each subsection, as seen below.

- ٧ -

Table (1): Size in million words of the spoken and written subsections of COCA:

Spoken	Written
256.690.399	746.199.355

Furthermore, the current study examined the use of these synonymous adjective in the eight genres (SPOKEN and TV/MOVIES, BLOG, WEB, FICTION, MAGAZINE, NEWSPAPER and ACADEMIC) of COCA separately. Qualitatively, the researchers followed the descriptive, narrative analytical style to find justification and suitable analysis for the behavior of the corpus data. It is important to mention that CLC represents only academic written texts, therefore a comparison between the academic genre of COCA and CLC is made.

4. Results and Discussions

Question 1: What are the similarities and differences among the four synonymous adjectives, namely small, little, petite, and tiny, in terms of frequency in COCA and CLC?

As mentioned before, per million frequency is used to show comparable analysis of results of the two corpora because they are different in size. Table (2) illustrates the frequencies, raw and per million of the four adjectives with many variations in their ranking with a slight similarity in COCA and CLC.

Table (2): Frequencies of the near synonyms small, little, tiny and petite in COCA and CLC

	COCA				CLC		
NO.	Adjective	Frequency	PM	NO.	Adjective	Frequency	PM
1.	Small	323442	325.71	1.	Small	1496	502.74
2.	Little	753793	759.09	2.	Little	1397	469.47
3.	Tiny	58060	58.47	3.	Tiny	67	22.52
4.	Petite	2302	2.32	4.	petite	0	0.00

* PM: Per Million

The data extracted from the Table elucidate that little and small have scored the highest ranks with the highest frequencies per million (PM) in COCA. In connection to the results, the behavior of the corpus shows that the adjective little has reached the highest frequency where it has scored (759.09) PM. The adjective small is the second adjective with a high frequency (325.71) PM compared to the others. The adjective petite has scored the lowest at about (2.32) PM. What is also noticeable is that the frequencies per million in COCA are higher for all adjectives. Concerning CLC results, the adjective small has the lion's share among all other adjectives with the frequencies of (502.74) PM. The adjective little has scored (469.47) PM, making it the second adjective with a higher frequency. Seemingly, the data show that tiny has a low frequency of (22.52) PM compared to the CLC adjectives. Surprisingly, petite has no frequencies. Based on COCA and CLC, the findings of both corpora revealed that the most frequent adjectives are small and little. Moreover, the findings also showed that tiny had scored the lowest frequencies in both corpora. It has been claimed that little and small have been used in the Old English words. The adjective little was used as a term of endearment since the 1560s. Around the eighteenth century, the adjective little was used much more than small, as shown in the N-gram (https://books.google.com/ngrams/graph?content=only+a+little+child,only+a+small+child&year_start=1800&year).

However, in the 19th century, the use of the adjective small has increased (https://www.eslbase.com/grammar/small-little and https://www.etymonline.com). This indicates that the use of adjectives may vary over time due to several reasons. After the end of the colonization, the Americans began to enhance their language in all its aspects. The adjective little has a wider range of meanings in the language.

The adjective little is used more metaphorically to suggest some sort of emotion or the idea of smallness, implies nothingness of a thing or the meaning of short while small usually refers mainly to the size of something which tends to be literally about size. Another reason is the fact that these two adjectives had the highest frequencies is that little is used as an adjective or adverb and small is used alone or when accompanied with another adverb or adjective. Using **word feature** from https://www.english-corpora.org/coca/ , provides several resources to define the meanings of each adjective such as https://www.merriam-webster.com/ , https://dictionary.cambridge.org/ , https://www.dictionary.com/ , https://www.lexico.com/ , https://www.google.com/ , and https://www.etymonline.com/ . All adjectives share the same meaning of being limited in number or amount, referring to the physical size. However, little conveys various meanings such as small in size, young, short time, unimportant something or someone, showing endearing feelings (cf. that's my little boy) or insulting (cf. what a little man). Therefore, little is more commonly used because it carries a lot of meanings.

Question 2: What are the similarities and differences among the four synonymous adjectives, namely small, little, petite and tiny, in terms of genres and grammatical patterns in COCA and CLC?

❖ Distributional analysis of the use of the four synonymous adjectives in the spoken and written subsections in COCA.

Table (3): Distributional analysis of use of the four synonymous adjectives in the spoken and written subsections in COCA

NO.	Adjective	Spoken		Written	
		Frequency	PM	Frequency	PM
1.	Small	316.08	123.14	2,305.93	309.02
2.	Little	2,346.76	914.24	3,706.77	496.75
3.	Tiny	68.03	26.5	403.82	54.12
4.	Petite	2.4	0.09	16.32	2.19

* PM: Per Million

Table (3) illustrates the raw and normalized frequency of use in COCA of the four adjectives in the written and the spoken forms. The normalized frequency of use of each adjective is calculated after determining the size of the two subsections of COCA (spoken and written see Table (1) in 3.3). Table (3) displays PM frequency of each target query, demonstrating that all adjectives except for little are more used in the written form, whereas it is more commonly used in the spoken subsection of COCA. However, little still records the highest frequency of use in the written subsection too. Again, we can assume that the various meanings of little might be the reason why speakers and writers of American English use it more frequently than small, tiny and petite.

- ٩ -

❖ **Genre analysis of the four synonymous adjectives in COCA**

Table (4): A descending order of the four synonymous adjectives cross genres in COCA in per million.

NO.	Adjective	Genre							
1.	Small	FICTION 512.78	MAG 486.39	NEW 343.31	ACAD 343.22	WEB-GEN 317.15	BLOG 303.08	SPOKEN 193.13	TV/MORE 122.95
2.	Little	TV/MOVIE 1.391.74	FICTION 1.082.67	SPOKEN 955.02	BLOG 618.94	WEB-GEN 597.91	MAG 583.46	NEWS 515.59	ACAD 308.2
3.	Tiny	FICTION 155.05	MAG 91.55	NEWS 53.79	BLOG 43.88	WEB-GEN 41.69	TV/MORE 37.04	SPOKEN 30.99	ACAD 17.86
4.	Petite	FICTION 5.57	MAG 4.08	NEWS 2.74	TV/MORE 1.5	ACAD 1.39	BLOG 1.37	WEB-GEN 1.18	SPOKEN 0.9

* ACAD= ACADEMIC   MAG= MAGAZINE

Table (4) shows the findings related to the order of the four adjectives in COCA according to their frequencies per million in different genres. It is apparent from this Table that the three adjectives small, tiny, petite scored the highest in FICTION. The tabulated data show that small, tiny, and petite scored (512.78 PM), (155.05 PM), (5.57 PM) respectively in COCA. A positive correlation was found between the adjective little and the genres: TV/MOVIE, FICTION, SPOKEN. Little is mainly used in TV/MOVIE recording (1,391.74 PM), then in FICTION recording (1,082.67 PM). This indicates that using it in daily life activities and imaginative settings due to its variety in meaning and usage in American English.

In contrast, the frequency of little is somehow limited in NEWS and ACADEMIC genres, recording 515.59 PM and 308.2 PM respectively, whereas small is least frequently used scoring (122.95 PM) in TV/MOVIE genre. This might reflect that little is not generally used in formal settings. Contrary to expectations, small, tiny, and petite recorded high in FICTION, MAGAZINE, and NEWS, which points out that they are more regularly used in the written genres of COCA and least used in the spoken genres of COCA (SPOKEN and TV/MOVIE). Another important finding is that tiny is least commonly used in ACADEMIC recording (17.86 PM). Even though the previous Table (4) illustrates the descending order of use of each adjective, it does not show a clear picture of the use of those adjectives' cross genres in COCA. As the chart shows, clearly, the use of little dominates all genres of COCA, except with a slight difference in use in ACADEMIC where small (343.22 PM) recorded a higher frequency than little (308.2 PM). Although small (512.78 PM) is mostly used in FICTION, little (1,082.67 PM) is twice as much used in FICTION than small.
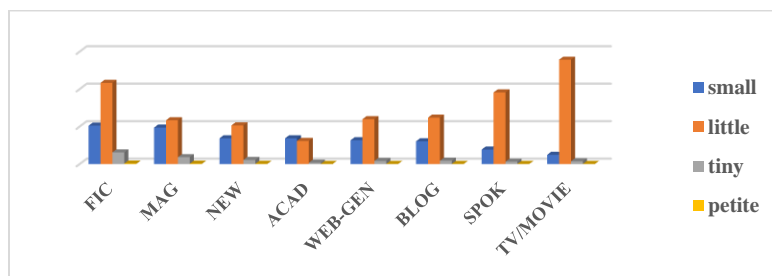

Chart (3): A comparison between the four synonymous adjectives cross genres of COCA

Examining the generated Concordances of the adjective revealed that this adjective is used differently in relation to parts of speech (POS): as a noun, an adjective, and an adverb. Therefore, the researchers examined the grammatical behavior of little to distinguish each POS of it:

1. **Little as a noun:** in this case little acts as a subject of the clause, usually proceeded by an article (specifically a) followed by a verb. In some cases, little is preceded by the article or zero article. It refers to a short distance, a small degree, amount, or particularly nothing, as seen in the following examples:

– his man will destroy what little is left of our U.S. Constitution, if he allowed to proceed without any resistance from the spineless and nauseating White elitist race traitors who comprise the current' establishment' of the despicable GOP. [COCA-BLOG]

– Using illusion to scare the masses has been going on since ancient times when rulers paid the scientists of their day to create steam pumps to force animal blood to flow from stone statues. Little has changed since then. [COCA-WEB]

2. **Little as an adverb that modifies an adjective or a verb**: meaning not much, slightly or not at all. For instance:

– A little known fact is that most Dutch Ph.D. programs do not involve tuition. [COCA-BLOG]

– Little does he know that I'm paying their best price. [COCA-MOVIE/TV]

3. **Little as an adjective:** modifying a noun referring to something or someone small in size or young, few numbers, not much, brief of short time or insignificant entity (an offense). For example:

– As a little girl growing up in an immigrant Chinese household in America, I never thought I was pretty. [COCA-BLOG]

– Fight it with your very lives people or the little people of the world will never again have a voice, or anything not deemed necessary by some government bureau! [COCA-BLOG]

– Even when laws are written and then enforced, there is little evidence that they readily translate into improvements for workers. [COCA-BLOG]

– As a' Claw/Crab Grip' user, it took me a little while to get used to Ragnarok Odyssey's camera, which can be controlled using the Vita's right analog stick. [COCA-BLOG]

– So, killing the story isn't going to protect your little man, Obama is history. [COCA-BLOG]

– Interestingly, searching the corpus showed that little as a noun is used in the plural form viz: littles to refer to small number of few members or group of young people. This plural form of little is only used in BLOG and WEB subsections of COCA. For example:

– What I have seen evolve in the littles with the most new and fresh talent is an interesting first splash. [COCA-WEB]

– Then we would run around being silly after the littles got tired and dropped off home. [COCA-WEB]

- ١١ -

Little is also used as a verb to show that something or someone is insignificant (belittle). As data search shows, little is the only adjective that can be formed to be used as a verb. Belittle is used 828 times in COCA (0.83 PM), whereas it was used only once in CLC. For instance,

- e.g. But even worse, people in this skeptical movement belittle those who want to increase the community feelings and welcoming nature and help our fellows. [COCA-WEB]

- e.g. Were the people aware of the variety of our social programmer, they would have undoubtedly restrained themselves from belittling our college. [CLC]

Such various uses of little adds another explanation why it records a higher frequency of use in COCA. Apparently, little and small can be interchangeable when they are used as adjectives and denotes size, for example:

- When I was little, I used to jump into puddles.

- When I was small, I used to eat a lot of chocolate.

On the other hand, Corpus data showed that small, tiny and petite could be used only as adjectives. Therefore, these adjectives are not interchangeable in several cases as 1. little conveys various senses than referring to the smallness of the physical size, 2. when used as a noun, an adverb or modifier. In the next section, a demonstration of comparative and superlative forms and the grammatical patterns of use of the four synonymous adjectives will be presented to show similarities and differences in use in COCA and CLC.

❖ A comparison between Academic of COCA and CLC:

Table (5): Per million frequencies of use of the synonymous adjectives in Academic of COCA and CLC

Adjective	Academic of COCA	CLC
Small	343.22	502.74
Little	308.2	469.47
Tiny	17.86	22.52
petite	1.38	0.00

Table (5) shows similarities in frequencies of use of the adjectives, small (343.22 in COCA and 502.74 in CLC) scoring the highest frequency in both corpora, followed by little (308.2 in COCA and 469.47 in CLC) then tiny (17.86 in COCA and 22.52 in CLC). However, the top three adjectives scored higher frequencies in CLC. In the next section, a grammatical description of the use of the four adjectives in both corpora.

Question 3: What are the similarities and differences among the four synonymous adjectives, namely small, little, petite, tiny, in terms of grammatical forms, collocations, and preceding verbs in COCA and CLC?

❖ Forms of the four synonymous adjectives in COCA and CLC

Table (6): Adjectives in comparative and superlative forms in COCA and CLC:

NO.	Adjective		COCA		CLC	
			Frequency	PM	Frequency	PM
1	Small	Smaller	57280	57.78	105	35.29
		Smallest	8258	8.32	11	3.69

No.	Word	Form		*PM		*PM
2.	Little	Littler	425	0.25	1	0.33
		More little	228	0.23	0	0.00
		Littlest	656	0.66	1	0.33
		Most little	38	0.04	0	0.00
3.	Tiny	Tinier	208	0.21	0	0.00
		Tiniest	1457	1.47	1	0.33

* PM: Per Million

The previous Table (6) displays the comparative and superlative forms of small, little, and tiny in COCA and CLC. It is important to mention that petite is not used in any form, whether in comparative or superlative, in both corpora, maybe because it is borrowed from French. As the findings demonstrate, little is formed with –er and more to form the comparative adjective. So, is there any difference in meaning and use between the two forms or are they the same?

The difference in meaning between littler and more little is that users of American English use the former form to compare between two entities in relation to size, amount, and age. Whereas they use the latter form to mean another thing (another question, wish, or job).

- He's Littler than me. We don't even wear the same size. [COCA-TV]

- who sat on my other side, gave me a tap on the shoulder and said Haven't you beard? And then, in a littler voice just for me, Her man's at a Magdalene laundry. [COCA-FIC]

- Dead man, dead man, come alive # Come alive by the count of five. # And in those long outside hours with my brother I went back to games I had not played in years, the games of a littler child, such as our father might have played before he left. [COCA-FIC]

- So, what was that all about [?] Just one more little Christmas wish to make . [COCA-MOVIE]

- I have one more little surprise for you. Pizza party part two. Oh, wow. Two pizzas. And toppings. [COCA-TV]

- I hope it's okay if I ask you one more little question. [COCA-MOV]

- Both forms, littlest and most little are also used to refer to size, age and amount. Moreover, most is used as a modifier of the adjective little to refer to most of a group, for example:

- I've had friends complement our littlest room in the house. [COCA-WEB]

- You need parents to start really watching their kids instead of being on their cell phone and tablets while the littlest kids go running off in every direction looking for "Mommy or Daddy". [COCA-WEB]

- You know I have received a letter from Nepal or in the most little town in India or in Chile. [COCA-WEB]

- However, Dana did not teach in the main, 21st century technical skills which is true of most little liberal arts colleges. [COCA-ACAD]

- Because while most little girls love horses, they don't have moms who are riders and who have space to keep horses and have money to afford it. [COCA-BLOG]

- ❖ Grammatical patterns used with the four synonymous adjectives in COCA and CLC

- ١٣ -

In this section, grammatical analysis of patterns of the four synonymous adjectives will be discussed to see whether both corpora show the same use of the grammatical structure in terms of nouns they modify, modifiers come before, and verbs occur before these adjectives. It has been claimed that synonymous adjectives share some aspects of the meaning but do not necessarily have the same grammatical structure (Liu 2010).

Table (7): grammatical patterns with the four synonymous adjectives in COCA and CLC

Grammatical structure	Small	little	Tiny	petite
A. With modified Ns	1.(det.) small +N 2.prep. + (det.) small+ N 3.(det.) + adj.+ small + N	1.(det.) little+ N 2. prep. + (det.) little+ N 3.(det.) + adj.+ little +N	1.(det.) tiny+ N 2. prep. + (det.) tiny+ N 3.(det.) + adj.+ tiny +N	1. (det.) petite + N 2. prep. + (det.) petite +N 3.(det.) + adj.+ petite +N
B. with verbs before the adjective	1.subj+be+ (det.) small 2.subj+ V + (det.) + (adj.) +small + N 3. subj.+ V+ obj.+ (det.) small	1.subj+ be+ (det.) little 2.subj+V +(det.) +(adj.) + little + N 3. subj.+ V+ obj.+ (det.) little	1.subj+be+ (det.) tiny 2.subj+V+(det.) + (adj.) +tiny+ N 3. subj.+ V+ obj.+ (det.) tiny	1.subj+be+ (det.) petite 2.subj+V+(det.) +(adj.) + petite + N 3.subj+V+ obj.+ (det.) petite
C. modifiers before the adjective	1.subj+be+ (det.) +adv+ small (N) 2.subj+be+ (adv) adj.+ prep.+ adv+ small+ N	1.subj+ be+ (det.) + adv+ little (N) 2.subj+be+ (adv) adj.+ prep.+ Adv +little + N	1.subj+be+ (det.) + adv+ tiny (N) 2.subj+be+ (adv.) adj.+ prep.+ adv+ tiny+ N	1.subj+be+ (det.) +adv + petite (N) 2.subj+be+ (adv.) adj.+ prep.+ adv+ petite + N

det. = determiner   subj. = subject   v = verb   obj. = object   adv. = adverb   prep.= preposition   art.= article

In accordance with the results, the grammatical patterns of the four synonymous adjectives used in COCA and CLC show that native and non-native users employ the same grammatical patterns with small. Table (7) demonstrates the grammatical patterns with the four synonyms small, little, tiny, and petite in both corpora in terms of nouns modified by the adjectives, verbs before the adjective and modifiers that occur before the adjective. First, the noun phrase that includes the target adjectives.

1. **Nouns modified by the adjective (NP):** all adjectives pre-modify a noun, which can be countable/uncountable, common/ proper. Interestingly, little is more commonly used with abstract nouns that reflect ideas, concepts and feelings. Small and tiny are rarely used with this type of noun, whereas petite never occurs with this type in COCA. In CLC, only little is used with abstract nouns.

2. **Determiners**: All the synonymous adjectives occur between a determiner and a noun. The determiner might be article/zero article (a/an), demonstrative (this, that, these, those), possessive pronoun (my, her, his, their, our), quantifier (few, some, every) and a number (one, two, three).

3. **Prepositional phrase (PP):** all adjectives can modify the object of a preposition (a noun), the adjective occurs between the preposition and a noun. The determiners mentioned above can occur before the adjective inside the PP.

   – I'm so happy they have a tiny size, which reduces the stress of having to finish the whole thing. [COCA-Blog]

– It was only a small mistake but it changed my life forever. [CLC -FCE]

– I have small information about his friend. [COCA-TV]

– I'm ashamed because I had a little accident and I broke your crystal vase. [CLC -FCE]

– Every morning the petite woman with long black hair walks Edith one long block to the

– the town's main street so the teenager can take a cab to her middle school, where she is studying.

– to be an electrician. [COCA-NEW]

– To walk into a slum and rescue this tiny baby. [COCA-Blog]

– He still had one ambition; he still had that last chance which he could use to turn his life around which was to see his little baby daughter walk again. [CLC-CPE]

– Funding and implementation aside, what matters now is that a design has been agreed on.

– and -- except for working out a few small design details -- the public process is over. [COCA-Blog]

– During your short trip you should bring some bottles of water and snacks although there

– plenty of little bars along the way. [CLC-CAE]

Second, the four synonymous adjectives occur as a subject complement (subj.+ be+ (det.) Adj.) or object complement (subj.+ v+ obj.+ (det.) Adj.) in COCA and CLC. It is important to mention that the most used verb with the examined adjectives in both corpora is the verb **be**. For example:

❖ Subject complement

– When I was little I was not very calm, I admit it, but this didn't mean that my parents had the right to beat me every time I did something wrong. [CLC-CPE]

– a small fraction of a regular Manhattan city block. Its imprint is very small. [COCA-BLOG]

❖ Object complements

– Embedding a Google Calendar is your sidebar is a good option if you want the calendar to be visible on every page of your blog, but it does make the calendar small. [COCA-WEB]

– We also find the clubroom very small and depressing and that is not very appealing to new members or to younger members who want more lively meetings. [CLC-CAE]

Third, small, little, tiny and petite can be modified by intensifiers such as very, too, so, really, rather and really in COCA and CLC. For example:

❖ Modifiers before the adjective

– In this plain and sunny place, you can rent a bike by payment of a very small amount from automatic bike rental spots that have been placed by the government all over the city. [CLC-CPE]

– ١٥ –

- For instance, the accommodation is not very good, the beds are too little and old. [CLC-FCE]

- For instance, really tiny motions or really fast motions, these are things that we don't typically see with our eyes. [COCA-SPOK]

- I know. She was very petite. Probably a size fou [COCA-FIC]

❖ Collocates with the four synonymous adjectives in COCA and CLC

In this section, a frequency analysis of the most common nouns modified with the four synonyms and verbs that occur before them is presented. To do that, collocates are lemmatized to show accurate results. Lemmatizing means to group all inflected forms, which share the same meaning, of a word so they can be analyzed as one word.

Table (8): Most frequent collocations, in raw frequency, with the four synonymous adjectives in COCA

No.	Small	Little	Tiny	Petite
1.	Business (13219)	Bit (46610)	Bit (1369)	Woman (120)
2.	Town (9048)	Girl (22822)	Fraction (853)	Command (44)
3.	Group (7932)	Boy (10999)	Town (484)	Frame (38)
4.	Number (5451)	While (6868)	Room (336)	Brunette (28)
5.	Amount (5359)	Time (6204)	Minority (314)	Bourgeoisie (19)

Table (9): Most frequent collocations, in raw frequency, with the four synonymous adjectives in CLC

No.	Small	Little	Tiny
1.	Shower (92)	Bit (216)	Minority (2)
2.	Town (74)	Girl (57)	Bite (2)
3.	Village (71)	Boy (39)	Suitcase (2)
4.	Group (43)	Child (30)	
5.	Room (35)	Money (29)	

It is apparent from table (8) and (9) that the collocations among these adjectives are not very similar. Small has got only two similar collocations, " Town" and "Group" while the little has got more similar ones such as " bit", girl", and "boy". In COCA, the adjective petite is used with other French words that match it. The reason behind this is that it has a French origin. Turning to CLC, another unexpected finding is that tiny had no similar collocations at all. The most surprising aspect of the data is that petite is not used among non-native learners. Second, the results show that the verb to be is the most used verb with the adjectives as mentioned earlier by native and non-native users of English. The following tables demonstrate the top three verbs before small, little, tiny and petite.

Table (10): top three verbs before the four synonymous adjectives in COCA and CLC (in raw frequency)

	COCA					CLC		
NO.	Small	Little	Tiny	Petite	NO.	Small	Little	Tiny
1.	Be (30998)	Be (107857)	Be (5346)	Be (304)	1.	Be (141)	Be (31)	Be (2)
2.	Make (2275)	Get (14027)	See (383)	Say (10)	2.	Become (8)	Know (3)	

3.	Start (998)	Take (5776)	Make (289)	See (8)	3.	Get (4)

This table is quite revealing in several ways. First, unlike the other verbs used before these adjectives, the verb to be is used significantly in corpora. Second, the verbs are very different in both corpora. What is more, the adjective petite is not used at all by learners of English. It could be concluded that learners might find it easier to express their ideas by using a simple basic verb in their sentences.

4.1 Conclusions

The purpose of the current study was to determine the similarities and differences among the four adjectives. The analysis of the results has shown that although these adjectives seem to be similar in isolation, some cases are not interchangeably used in all contexts. The most significant findings from this study are as follows:

1. Based on COCA and CLC, the general results have shown that the most frequent adjectives were small and little. The results of both corpora have also indicated that tiny has scored a low frequency. Surprisingly, the frequencies of petite elucidated that it is rarely in COCA and not used at all in the CLC.

2. The results of COCA have clarified a strong tendency for using the little in TV/ Movie, fiction, and spoken genres. Moreover, there is another tendency for Americans to avoid using these adjectives in news and academic texts. What is also noticeable is that small, little, and petite are mainly used in more formal settings such as news and magazines. Since CLC includes only academic written texts, it was compared with the frequency with the Academic subsection of COCA. The findings demonstrated the similarity of frequency of use between native and nonnative users of English that small (343.22 in COCA and 502.74 in CLC) scoring the highest frequency in both corpora, followed by little (308.2 in COCA and 469.47 in CLC) then tiny (17.86 in COCA and 22.52 in CLC).

3. In both corpora, the grammatical patterns of these adjectives have revealed remarkable similarity except for little. A different pattern is seen with little+ verb to be in COCA. Hence, this structure is not found in CLC. Little is used as the subject of the clause followed by a verb, usually the verb to be.

4. Another striking result concerning the four adjectives comparison is that petite does not have comparison forms in COCA and CLC. Furthermore, little has two comparison forms that vary in meaning.

5. The findings of COCA and CLC reveal that all adjectives tend to use the verb to behave similarly with before them.

6. These adjectives are not interchangeable in several cases as when 1. little conveys various senses than referring to the smallness of the physical size, 2. when it is used as a noun or an adverb.

5. Recommendations

Based on the obtained results in the current paper, the researchers propose the following recommendations:

1. Researchers may conduct a further corpus-based study on the same adjectives dealing with various aspects such as meaning, idiomatic expressions, and colligation of these adjectives.

- ١٧ -

2. Further comparative studies could be conducted using the same adjectives but different corpora in different languages such as Arabic or French corpora.

3. English language teaching may use corpora as a helpful language teaching material. This goal could be achieved by enhancing the students' ability to use the corpora as a practical learning tool to detect the differences between words in a set of synonyms autonomously. Using corpora can facilitate learners' vocabulary acquisition.

4. Corpora data has always been a vital source for pedagogical implications that helps curricula planners to improve them in showing fundamental differences among synonyms in English teaching books.

Conflicts of interest statements

All authors declare that they have no conflicts of interest.

References

Bergdahl, M. (2009). *Is he beautiful, handsome or good-looking: A study of putative synonymy in three adjectives* [Unpublished Master's Thesis]. Lulea University of Technology, Lulea, Sweden.

Cai, J. (2012). *Is it "Great" Enough? A corpus-based Study of "Great" and its Near Synonyms* [Unpublished Master's Thesis]. Ball State University, Muncie, Indiana.

Chanchotphattha, K. (2013). *A Corpus-based Study of English Synonyms: Different, Diverse, and Various* [Unpublished Master's Thesis]. Thammasat University, Bangkok, Thailand.

Chung, S. (2011). A corpus-based analysis of "create" and "produce". *Chang Gung Journal of Humanities and Social Sciences, 4*(2), 399–425.

Divjak, D. (2006). Ways of intending: Delineating and structuring near synonyms. In *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (pp. 19–56).

Davies, M. (2009). The 385+ million-word corpus of contemporary American English (1990–2008) design, architecture, and linguistic insights. *International Journal of Corpus Linguistics, 14*(7), 159–190.

Divjak, D., & Gries, S. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory, 2*(1), 23–60.

DiMarco, C., Hirst, G., & Stede, M. (1993, March). The semantic and stylistic differentiation of synonyms and near-synonyms. In *AAAI Spring Symposium on Building Lexicons for Machine Translation*, 1, 114–121.

Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics, 28*(2), 105–144.

Eslbase. *Small & Little.* Retrieved from https://www.eslbase.com/grammar/small-little

Etymonline. *little (adj.).* Retrieved from https://www.etymonline.com/word/little

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

Hoffmann, K. (2014). *A corpus-based analysis of the near synonyms nice, kind, lovely, friendly, gorgeous and pleasant.* Retrieved 16 Nov 2017, from https://www.academia.edu/7541436/A_corpusbased_analysis_of_the_nearsynonyms_nice_kind_lovely_friendly_gorgeous_and_pleasant

koshal. (2011). *Difference Between Small and Little.* Retrieved 12 Mar 2021, from https://www.differencebetween.com/difference-between-small-and-vs-little/

Laurea, T. (2014). *Different Forms, Different Meanings? A Lexicographic and Corpus-based Study of Four Sets of English Near-synonyms* [Unpublished Master's Thesis]. Università degli Studi di Padova.

Lee, C., & Liu, J. (2009). Effects of collocation information on learning lexical semantics for near synonym distinction. *Computational Linguistics and Chinese Language Processing, 14*(2), 205–220.

Liu, D., & Espino, M. (2012). Actually, genuinely, really, and truly: A corpus-based behavioral profile study of near-synonymous adverbs. *International Journal of Corpus Linguistics, 17*(2), 198–228. http://dx.doi.org/10.1075/ijcl.17.2.03liu

Liu, D. (2010). A corpus-based behavioural study of the near-synonyms. *International Journal of Corpus Linguistics, 15*(1), 56–87.

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

Meyer, C. (2002). *English corpus linguistics: An introduction.* Cambridge: Cambridge University Press.

Morimoto, A. (2020). Using corpora to teach the differences between reservation and appointment. *TESOL Working Paper Series, 18*, 111–125.

Murphy, L. (2010). *Lexical meaning.* London: Cambridge University Press.

Nisani, F. (2015). *A Corpus-based Study of English Synonyms: Possible, Probable, and Likely* [Unpublished Master's Thesis]. Thammasat University, Bangkok, Thailand.

Petcharat, N., & Phoocharoensil, S. (2017). A corpus-based study of English synonyms: Appropriate, proper, and suitable. *LEARN Journal, 10*(2), 9–24.

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

Putri, W., Rajeg, M., & Wandia, K. (2017). The study of adjective admirable and its near synonyms in Corpus of Contemporary American English (COCA). *Jurnal Humani, Fakultas Ilmu Budaya Unud, 18*(2), 144–152.

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

Sinclair, J. (2005). Corpus and text—basic principles. In M. Wynne (Ed.), *Developing Linguistic Corpora: A Guide to Good Practice.* New York: Routledge.

Sirikan, S. (2015). *A Corpus-based Study of English Synonym: Appalling, Horrific, Horrible, and Hideous* [Unpublished Master's Thesis]. Thammasat University, Bangkok, Thailand.

Webb, S. (2007). The effect of synonym on second-language vocabulary. *Reading in a Foreign Language, 19*, 120–136.

Su, Q., & Zhan, W. (2020). From minimal contrast to meaning construct (corpus-based, near synonym-driven approaches to Chinese lexical semantics). *Springer Singapore.* https://doi.org/10.1007/978-981-32-9240-6

Supachal, S. (2015). *A Corpus-based Study of the English Synonyms: Deadly, Lethal, and Fatal* [Unpublished Master's Thesis]. Thammasat University, Bangkok, Thailand.

Taylor, J. (2003). Near synonyms as co-extensive categories: "high" and "tall" revisited. *Language Sciences, 25*(3), 263–284.

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

Webb, S. (2007). The effect of synonym on second-language vocabulary. *Reading in a Foreign Language, 19*, 120–136.

Yang, B. (2016). A corpus-based comparative study of "learn" and "acquire". *English Language Teaching, 9*(1), 209–220. https://doi.org/10.5539/elt.v9n1p209