LEARN Journal: Language Education and Acquisition Research Network
ISSN: 2630-0672 (Print) | ISSN: 2672-9431 (Online)
Volume: 16, No: 1, January – June 2023

Language Institute, Thammasat University
https://so04.tci-thaijo.org/index.php/LEARN/index

Analysis of Collocations and Semantic Preference of the Near-synonyms: *Blank*, *Empty*, and *Vacant*

Pallapa Lertcharoenwanich

pallapa.lert@bru.ac.th, Business English Department, Buriram Rajabhat University, Thailand

APA Citation:
Lertcharoenwanich, P. (2023). Analysis of collocations and semantic preference of the near-synonyms: Blank, empty, and vacant. *LEARN Journal: Language Education and Acquisition Research Network, 16(*1), 365-383.

Received
03/11/2022

Received in revised form
26/11/2022

Accepted
03/12/2022

ABSTRACT

For English language learners, mastering the use of near-synonyms can be challenging. Despite the semantic similarities of English synonyms, they are not interchangeable in all contexts. The objectives of this corpus-based study are to examine differences between the near-synonymous adjectives *blank, empty* and *vacant* based on the degree of formality from their distribution across genres and to analyze the noun collocates in relation to semantic preference drawn from the Corpus of Contemporary American English (COCA). The findings indicate that these synonyms are uncommon in formal English, i.e., academic texts and they tend to have quite different occurrences and distributions across genres (i.e., *blank* and *empty* most frequently occur in fiction whereas *vacant* mostly occurs in newspaper) despite a similar degree of formality. For the noun collocates and semantic preference, these three adjectives have a weak near-synonymous status as they share only one noun collocate and one similar theme of semantic preference. Thus, despite their similarity in meanings, these synonyms co-occur with specific noun collocates in a particular context.

Keywords: collocations, semantic preference, near-synonyms, COCA

Introduction

Understanding lexical semantics, particularly synonyms or words with analogous meanings, is regarded to be crucial yet difficult for non-native English learners since each synonym has a unique meaning that is not entirely interchangeable (Edmonds & Hirst, 2002). Synonyms allow language users to choose one term over another with comparable meanings to express nuance to their writing or speaking. However, since English vocabulary has so many synonyms, mastering its usage may be challenging (Laufer, 1990). Apart from their meanings, dialects, degrees of formality (Jackson & Amvela, 2007) and grammatical patterns (Phoocharoensil, 2010) also have a role in the ability to discriminate between synonyms. Therefore, it is crucial for second language (L2) learners to examine words that commonly co-occur with synonyms in order to discover similarities and variations in their meanings and be able to communicate clearly and properly (Edmonds & Hirst, 2002).

In this study, the genres, collocations, and semantic preferences of the three synonymous adjectives *blank*, *empty* and *vacant* are examined. The selection of the target words is based on the notion that many L2 learners may misuse the target words as they are presented in dictionaries in a way that implies their interchangeability. Additionally, even if lexical collocations may be offered, the genres and semantic preferences are typically absent or just obliquely conveyed. Thus, the major goal of this present analysis is to further distinguish the target synonyms using data from a large corpus of American English, known as the Corpus of Contemporary American English (COCA). Prior to considering COCA data, the definitions and applications of the target synonyms from online learner dictionaries are compared to provide a comprehensive picture of how the terms are used, namely Longman Dictionary of Contemporary English, Oxford Advanced Learners' Dictionary, and Cambridge Dictionary as illustrated in Table 1.

Table 1

Definitions and Examples of Blank, Empty and Vacant from the Online Dictionaries

	Longman Dictionary of Contemporary English	Oxford Advanced Learner's Dictionary	Cambridge Dictionary
1. *blank*	**1**. without any writing, print, or recorded sound	**1.** empty, with nothing written, printed or recorded on it	**1.** empty or clear, or containing no information or mark

	e.g., *In the* **blank** *spaces at the bottom of the chart, add other words you need.* **2.** a blank face or look shows no emotion, understanding, or interest e.g*., Zoe looked at me with a* **blank** *expression.*	e.g., *She turned to a* **blank** *page in her notebook* **2.** showing no feeling, understanding or interest e.g., *Suddenly my mind went* **blank.**	e.g., *a* **blank** *sheet of paper* **2.** showing no understanding or no emotion in the expression on your face e.g., *a* **blank** *stare/ expression* **3.** complete and absolute e.g., *a* **blank** *refusal/ denial*
2. *empty*	**1.** Having nothing inside e.g., *The fuel tank's almost* **empty**. **2.** An empty place does not have any people in it e.g., *The hall was* **half-empty**. **3**. not being used by anyone e.g., *I spotted an* **empty** *table in the corner.* **4.** unhappy because nothing in your life seems interesting or important e.g., *The divorce left him* **feeling empty** *and bitter*	**1.** with no people or things inside e.g., *I noticed an* **empty** *space on the bookshelf.* **2.** (of something that somebody says or does) with no meaning; not meaning what is said e.g., *an* **empty** *promise* **3.** (of a person, or a person's life) unhappy because life does not seem to have a purpose, usually after something sad has happened e.g., *My life seems* **empty** *without you.*	**1.** not containing any things or people e.g., Shall I take the **empty** bottles for recycling? **2.** not sincere or without any real meaning e.g., *They're just* **empty** *promises.* **3.** without purpose or interest e.g., *I felt* **empty***, like a part of me had died.*
3. *vacant*	**1**. a vacant seat, building, room, or piece of land is empty and avai lable for someone to use e.g., *Only a few apartments were still* **vacant**. **2.** a job or position in an organization that is vacant is available for someone to start doing e.g., *He was offered the position of headmaster when it fell* **vacant**. **3.** an expression that shows that someone does not seem to be thinking about anything e.g*., He was looking round with a* **vacant** *look on his face.*	**1.** (of a seat, hotel room, house, etc.) empty; not being used e.g., *There are very few* **vacant** *properties available in the area.* **2.** *(formal)* if a job in a company is vacant, nobody is doing it and it is available for somebody to take e.g., *The position left* **vacant** *in July has not yet been filled.* **3.** showing no sign that the person is thinking of anything e.g., *She had large* **vacant** *brown eyes.*	**1.** not filled or occupied; available to be used e.g., *The hospital has no* **vacant** *beds.* **2.** A vacant job is one that no one is doing and is therefore available for someone new to do e.g., *The position fell* **vacant** *when Rodman was promoted.* **3.** showing no interest or mental activity e.g., *She had a* **vacant** *look/expression on her face.*

Table 1 illustrates that the three target adjectives are near-synonyms as they share similar core meaning of 'absence of content, people, things or emotion.' However, despite being synonymous in certain situations, they cannot always be used in the same context. For example, these three synonyms can be used to express the meaning of 'showing no interest or emotion,' whereas *empty* and *vacant,* but not *blank,* are commonly used with places. Thus, rather than their intended meaning, the issue of collocation should be taken into consideration as the target words all have quite distinct meanings when collocated with particular words. According to Szudarski (2018), L2 English learners who have not had enough exposure to the language may believe that the terms have identical meanings, which will cause their L2 usage to stray from the language's natural usage. In addition, Longman Dictionary of Contemporary English (Online) suggests that *blank* and *vacant* are categorized as medium-frequency core vocabulary while *empty* is considered to be of high-frequency. Thus, it is worthwhile to examine their usage and common collocations in order for learners to use them correctly and appropriately.

Literature Review

Since this paper focuses on the analysis of semantic preference and collocations of the target synonyms *blank, empty* and *vacant,* these concepts will be elaborated in more detail in this section.

Synonymy

A crucial yet complex language aspect in the study of lexical semantics is synonymy, often known as semantic equivalence. Despite their similarity, synonyms have different meanings, implications, and registers (DiMarco et al., 1993; Norris, 2016). Due to historical reasons, English is particularly rich in synonyms because of the influences from many other languages; thus, it contains pairings of local and foreign terms that have the same meanings (Palmer, 1981). As the appropriate use of synonyms allows language users to convey precise and effective meaning in the right context (Laufer, 1990; Liu & Espino, 2012), it is considered to be one of the challenges learners have while acquiring vocabulary. There are two main types of synonyms: absolute synonyms and near-synonyms (Cruse, 1986; Edmonds & Hirst, 2002; Murphy, 2010).

Absolute synonyms, often referred to as perfect synonyms, are words having the exact same meaning that may be used interchangeably in all circumstances without changing the meaning, connotation, or style of the

sentence (Cruse, 1986; Edmonds & Hirst, 2002). This kind of synonym is quite uncommon and very rare, though, as it might result in redundancy of the language use (Phoocharoensil, 2020). According to Martinet (1955, as cited in Vicentini, 2003), the two words with the same meaning cannot co-exist for a long time in one language. This lexical adaptation specifically takes place due to the need for clarity and accuracy in communication. In addition, Clark (1992) further asserts that if two words have the same meaning, one of them will either disappear or take on a different meaning.

Near-synonyms, on the other hand, are words that are highly close in meaning but cannot be utilized interchangeably (Liu, 2013), mostly because of their subtle differences in meanings, collocations, or registers (DiMarco et al., 1993; Jackson & Amvela, 2007). Near-synonyms are much more common than absolute synonyms in language. An example of these kinds of synonyms are the adjectives *beautiful*, *attractive*, *lovely*, and *charming*, which are nearly synonymous and have corresponding conceptual or denotational meanings but are probably distinct in collocational terms (Xiao & McEnery, 2006).

In conclusion, it can be said that near-synonyms are more common than absolute synonyms. The meaning, style, and context of the language used may be impacted by the improper use of synonyms. As a result, it is crucial for the language learners to be able to employ near-synonyms appropriately by taking into account the criteria for differentiating near-synonyms, including, collocations and semantic preference which will be discussed in the next section.

Collocations and Semantic Preference

In order to distinguish between synonymous words, collocations and semantic preferences are often examined together as they are closely related (Selmistraitis, 2020; Szudarski, 2018). The significance of lexical associations, denoted by collocations, was perhaps first identified by Firth (1957), who emphasized that the meaning of a word is based not only on the word itself but also on how it is coupled with the words surrounding it. According to Lewis (2000) and O'Dell and McCarthy (2008), a collocation is the co-occurrence of two or more words. It is a combination of two lexical terms that commonly appear in close proximity to one another, as Timmis (2015) further underlined. This may affect how individuals choose English words in productive skills (i.e., writing and speaking) (Lindquist, 2009). Thus, for English language learners to be able to generate natural and standard English, sufficient collocational knowledge is required since the notion of collocation is seen as being fundamental in lexicology and vocabulary teaching and learning (Szudarski, 2018). Examples of the discrepancies in the collocational patterns of the synonyms *tasks* and *jobs* are given by Edmonds and Hirst

(2002) in which one can confront a *daunting task* but not a *daunting job*. This idea of collocational patterns is connected to other criteria in distinguishing the near-synonyms, that is semantic preference.

Semantic preference is described by Cheng (2012) as the propensity of lexical items that are restricted in specific semantic circumstances. Although collocational patterns are concerned with lexical constraint, the words' occurrences are determined by their semantic contexts (Flowerdew, 2012). It is also known as "the frequent co-occurrence of a lexical item with items expressing a particular evaluative meaning" (Hunston, 2007, p. 266). In other words, semantic preference is seen as a semantic setting where vocabulary terms frequently appear. For instance, Edmonds and Hirst (2002) note that the synonymous verbs *die* and *pass away* have different semantic preferences that *pass away* is used solely with humans and not with animals or plants. Another example elaborated by Partington (2004) on the findings about the semantic preference of the word *cause*. When *cause* is followed by a single object, it is typically associated with a sickness, such as *cancer* or *heart disease*; however, when *cause* is followed by two objects, the second object frequently indicates an unpleasant feeling, such as the word *comfort* in *causes them discomfort*.

As a result, semantic preference is closely related to collocations. In determining the appropriate use of lexical items, collocational patterns in a certain semantic context should be taken into account. Thus, in this study, linguistic features of the near-synonyms *blank*, *empty* and *vacant*, namely genres, word frequency, typical noun collocates as well as semantic preference of the noun collocates of each target words are investigated and analyzed to identify their differences by using a corpus-based data.

Previous Studies

Studies on synonym differentiation have been conducted using data from linguistic corpora, such as the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA). This corpus-informed data gives evidence of a word in addition to its dictionary definition, for example, degree of formality, connotations, and collocational patterns.

To begin with, Gu (2017) compared the synonyms *obtain* and *gain* in terms of semantic prosody, collocation, colligation, and genre using BNC, and Sketch Engine. It was found that the passive pattern with preposition was a frequent colligation for *obtain*. While abstract nouns were considerably discovered to collocate with *gain*, nouns were found to collocate with *obtain* most frequently to signify procedures or the importance of law and commerce. Consequently, the study discovered that *gain* had positive semantic prosody; however, *obtain* had neutral and mixed semantic prosody.

In a similar vein, Li (2019) summarized the semantic preference of the synonyms *preserve* and *conserve* by looking at the collocations using Sinclair's (1996) concept of extended lexical units (ELUs) and the data drawn from BNC. It was found that the frequency of *preserve* is more than six times greater than that of *conserve* which means that *preserve* is significantly more common than *conserve*. In terms of semantic preference, *preserve* implies to "maintain something as it is, without making any modifications," and has the semantic preference of something that is abstract yet significant, such as *integrity*, *anonymity*, *unity*, and *independence*. Contrarily, *conserve* means to "use as little of anything as possible so that it lasts a long time" and has the semantic preference of something relating to ecology, natural resources, and the environment, such as *moisture*, *biodiversity*, *energy*, and *heat*.

In the academic texts in COCA, Selmistraitis (2020) looked at semantic preference, semantic prosody, and distribution across nine academic genres of three pairs of synonymous adjectives: *succinct* and *concise*, *coherent* and *cohesive*, and *precise* and *accurate*. These near-synonyms were discovered to have various semantic preferences; nonetheless, *succinct* and *concise* had more semantic similarities than the other two pairs of synonyms. Additionally, the distribution of these three pairs of synonyms varies between academic genres, showing that they are not interchangeable in all situations. In addition, another recent corpus-based study conducted by Phoocharoensil and Kanokpermpoon (2021) uses COCA to analyze the genres and collocations of the two synonymous verbs *increase* and *rise*. The results show that both synonyms are often used in formal written genres. While the majority of the adverb collocates are shared by the target synonyms, there appears to be a noticeable difference in the nouns that each collocates with. *Rise* differs from *increase* in that it means "go higher" when used with nouns denoting the direction of a natural phenomenon, like the moon, the sun, or the tide.

More recently, Kruawong and Phoocharoensil (2022) examined the distribution of the three synonymous verbs: *teach*, *educate*, and *instruct* across genres and collocational patterns based on the data taken from COCA. The results of this analysis showed that, among the eight genres, *teach* was used the most frequently, and it was significantly more extensively and frequently used than *educate* and *instruct*. The frequencies across genres also showed that all the three synonyms occur more often in formal genres than spoken genre. The researchers suggest that categorization of noun collocates provides more informative data about each target synonym's concurrent authentic use which are beneficial for English language learners.

According to the previous studies, the efficiency of corpus linguistics as a method for differentiating synonyms has been supported extensively by past research (e.g., Gu, 2017; Li, 2019; Phoocharoensil & Kanokpermpoon, 2021). The distribution across genres, collocations, and semantic preference

could be employed to distinguish near-synonyms based on the use of language corpora. By collecting data from the Corpus of Contemporary American English (COCA), this study applies those criteria to examine the target near-synonymous adjectives *blank*, *empty*, and *vacant* in order to investigate their common occurrences and differences in greater depth. As a result, the following research questions are addressed:

1. What are the differences among the synonyms *blank*, *empty* and *vacant* across genres in relation to their frequencies of occurrence?

2. What are noun collocates in relation to semantic preference of the synonyms *blank*, *empty* and *vacant*?

Methodology

In this study, the near synonymous adjectives *blank*, *empty* and *vacant* were examined. The online version of three dictionaries: Longman Dictionary of Contemporary English, Oxford Advanced Learners' Dictionary, and Cambridge Dictionary provided the fundamental definitions of the target terms. The three target synonyms were further studied using corpus-based data from the Corpus of Contemporary American English (COCA). COCA was used as a data source in this investigation for three reasons. First, the data to be studied are made generalizable and reliable by COCA, which is known as the biggest corpus to include authentic American English. Second, COCA is regarded as a monitor corpus, with updated data being normally added each year to keep the corpus current. From 1990 to 2020, more than 25 million words are gathered annually (Davies, 2020). Lastly, COCA gathers information evenly throughout eight various genres (i.e., spoken language, fiction, popular magazines, newspapers, academic texts, TV and movie subtitles, blogs, and online webpages). The distribution of the target terms across genres in COCA has thus been comprehensively and thoroughly investigated.

To distinguish the target synonyms, i.e., *blank*, *empty*, and *vacant,* this present study concentrated on genre distribution, collocations, and semantic preferences that can be obtained from COCA. Since the target terms are adjectives, attention was given to the most common noun collocates. The frequency and Mutual Information (MI) score were taken into consideration while choosing these collocates. The MI value, according to Gablasova et al. (2017), was utilized to denote the strength of the co-occurrence between the target word and the collocation. However, since the MI score frequently assigns a high value to the collocations with low frequency, the MI score alone cannot yield accurate estimates of collocational strength (Cheng, 2012). In other words, collocations with high MI scores might be low in frequencies, making them unsuitable to be used in the analysis since they might not be a

good indicator of strong collocations. Thus, it is important to note that, in order to confirm the co-occurrence strength between the target words and the collocation, Schmitt (2010) advises using the MI score with caution and examining it along with frequency. In order to do this, the top thirty noun collocates of the target synonyms with the highest frequencies and the significant MI score level (≥3) (Cheng, 2012) were taken from COCA. Following that, the list of common noun collocates was scrutinized to determine their semantic preference. In other words, similar collocations were grouped under a common theme. Therefore, a clear picture of the distinctions between the target synonyms in terms of collocational patterns and context of occurrence was provided by this semantic categorization.

Results and Discussion

Frequency and Distribution of the Synonyms across Genres

In response to Research Question One, the formality of the target synonyms based on their overall frequencies and distributions across various genres were investigated as illustrated in the following table.

Table 2

Frequency and Distribution of the Synonyms across Genres

Blank			**Empty**			**Vacant**		
Genre	*Frequency*	*Per million (w/m)*	*Genre*	*Frequency*	*Per million (w/m)*	*Genre*	*Frequency*	*Per million (w/m)*
Fiction	3,917	33.10	Fiction	21,358	180.51	Newspaper	1,724	14.16
Webpage	1,843	14.83	Webpage	5,683	45.74	Fiction	1,417	11.98
Magazine	1,620	12.85	TV/Movie subtitles	5,789	45.20	Magazine	623	4.94
Blogs	1,573	12.23	Magazine	5,405	42.87	Webpage	609	4.90
Spoken	1,495	11.85	Newspaper	4,610	37.87	Blogs	525	4.08
Academic texts	1,268	10.59	Blogs	4,723	36.72	Academic texts	411	3.43
Newspaper	1,212	9.96	Academic texts	2,651	22.13	Spoken	356	2.82
TV/Movie subtitles	1,142	8.92	Spoken	2,743	21.75	TV/Movie subtitles	303	2.37
Total	14,070	14.17	Total	52,962	53.33	Total	5,968	6.01

It can be seen from Table 2 that, in terms of frequency and normalized frequency (Per million), *empty* (52,962 tokens or 53.33 w/m) occurs with the highest frequency among the three target synonyms followed by *blank* (14,070 tokens or 14.17 w/m) and *vacant* (5,968 tokens or 6.01 w/m).

In terms of genres, both *blank* (3,917 tokens or 33.10 w/m) and *empty* (21,358 tokens or 180.50 w/m) are the most frequent in fiction whereas *vacant* (1,724 tokens or 14.16 w/m) mostly occurs in newspaper. This is probably because the contents that appear in newspapers involve issues concerning occupational positions or advertisements about jobs or space vacancy. In addition, it is noticed that the frequency of *empty* (5,789 tokens or 45.20 w/m) is prominently higher than that of *blank* (1,142 tokens or 8.92 w/m) and *vacant* (303 tokens or 2.37 w/m) in TV and movie subtitles. Additionally, *empty* is least frequent in spoken (2,743 tokens or 21.75 w/m), whereas *blank* (1,142 tokens or 8.92 w/m) and *vacant* (303 tokens or 2.37 w/m) are least frequent in TV and movie subtitles.

Furthermore, it is evident that these synonyms are uncommon in formal English, i.e., academic texts since they are used in this genre at the bottom three lowest frequencies. This indicates that the three target synonyms have quite similar degrees of formality despite their different occurrence across genres which is in line with previous studies (Cai, 2012; Li, 2019; Petcharat & Phoocharoensil, 2017; Phoocharoensil, 2020; Phoocharoensil & Kanokpermpoon, 2021). In addition, it can be seen that empty is least frequently used in spoken context (2,743 tokens or 21.75 w/m). This is in contrast to the data from the dictionary which suggest that empty is listed as among the top 2,000 spoken words (Longman Dictionary of Contemporary English Online). This contradiction underlines the importance of employing corpus-based data when teaching English vocabulary.

Common Collocations: Analysis of Noun Collocates

To answer Research Question Two, this section analyzes the common noun collocates and their semantic preferences of the synonyms *blank*, *empty* and *vacant*.

Table 3

Common Collocations: Analysis of Noun Collocates

	Empty			*Blank*			*Vacant*		
	Noun collocates	Frequency	MI Value	Noun collocates	Frequency	MI Value	Noun collocates	Frequency	MI Value
1	space	1898	4.75	page	640	5.42	lot	578	4.78
2	room	1738	3.61	face	603	4.55	building	338	5.42
3	bottle	1273	6.09	check	551	6.82	seat	297	6.00
4	seat	1155	4.90	slate	484	9.28	house	271	3.39
5	chair	1039	5.02	space	451	4.72	position	270	4.99
6	glass	879	4.54	screen	400	5.79	land	260	5.28

7	street	861	3.15	stare	385	9.64	eye	216	3.69
8	stomach	677	5.87	paper	374	4.55	space	200	4.55
9	box	643	3.85	mind	352	4.04	home	187	3.39
10	bed	530	3.38	wall	336	4.60	property	156	4.94
11	can	413	5.48	eye	296	3.14	apartment	120	5.24
12	promise	391	4.49	canvas	292	7.91	office	111	3.16
13	apartment	390	3.89	look	266	4.10	street	105	3.17
14	cup	365	3.06	expression	215	5.26	storefront	76	9.21
15	bag	358	3.31	sheet	202	5.62	unit	73	4.17
16	suit	358	4.02	spot	174	4.60	store	58	3.24
17	beer	326	4.03	piece	140	3.13	spot	54	3.92
18	parking	284	4.40	range	105	3.61	expression	49	4.13
19	tomb	275	6.45	map	95	4.08	chair	46	3.58
20	nest	259	5.65	verse	92	6.04	housing	45	4.73
21	plate	257	3.66	card	91	3.05	title	41	3.38
22	tank	247	4.02	tape	80	4.11	block	38	3.44
23	shelf	244	4.81	ballot	64	4.64	acre	34	4.91
24	shell	244	4.81	fill	47	7.65	neighbor hood	29	3.05
25	sky	225	3.02	notebook	30	4.74	shop	24	3.03
26	container	220	4.92	pad	29	4.34	throne	23	5.26
27	bowl	218	3.18	gaze	28	3.98	parking	22	3.76
28	string	199	4.11	monitor	27	3.78	warehouse	20	5.25
29	desk	193	3.39	disk	24	4.14	desk	19	3.10
30	calorie	183	3.55	disc	22	4.44	parcel	19	6.43

Table 3 presents the top-30 frequencies and MI score (≥3) of the noun collocates that commonly occur with the synonymous adjectives *blank*, *empty* and *vacant*. It is evident that one noun collocate i.e., *space* is shared among the three near-synonyms. It can be concluded that these adjectives are not strongly synonymous. However, the data from the corpus indicate the near-synonymous status of *empty* and *vacant* as they share several overlapping noun-collocates, i.e., *seat, chair, street, apartment,* and *parking,* whereas there is no similar noun collocates other than *space* which occur between *empty* and *blank*. This shows a more closely related status of *empty* and *vacant* because, according to Szudarski (2018), words with similar meanings frequently have collocates in common. Furthermore, there are three nouns co-occurring between *blank* and *vacant*, namely *eye*, *spot*, and *expression* which indicates that *blank* and *vacant* are not quite strongly synonymous.

Nevertheless, the collocational data shown in Table 3 should be carefully examined. Due to their low frequencies or low MI score in COCA, several nouns that may be used with *blank*, *empty* and *vacant* are not featured in the table. For example, the word *building* can be collocated with both *empty* and *vacant* based on the data from COCA, but it does not appear to meet the study's selection criteria because its MI score as a collocate of *empty* (2.92) is

lower than 3, despite the words' high frequency (495 tokens) as co-occurring partners (338 tokens).

In the following section, noun collocates of the synonymous adjectives *blank*, *empty* and *vacant* will subsequently be grouped according to their semantic preferences. The list of collocations, as shown in Table 3, provides information that enables identification of the variety of lexical item linkages as well as the meaning relationships among collocates (Crawford & Csomay, 2016).

Table 4

Semantic Preference of Noun Collocates of Blank

1. WRITING AREA	ballot, canvas, card, check, fill, map, notebook, page, pad, paper, sheet, slate, tape, verse
2. AREA	piece, range, space, spot, wall
3. ELECTRONIC DEVICES	disc, disk, monitor, screen
4. STATE OF MIND	expression, eye, face, gaze, look, mind, stare

Table 4 shows that there are four main themes of the noun collocates of *blank* based on their semantic preferences. The majority of the noun collocates belongs to the first theme, WRITING AREA, which indicates 'the space to be written on', e.g.*, page, check,* and *slate,* as exemplified in (1) below. The next theme, AREA, is closely related to the first theme in which it expresses 'the space to be used or filled out', e.g., *space, piece,* and *wall* as in (2). The third theme is ELECTRONIC DEVICES which incorporates nouns related to 'a device that accomplishes its purpose electronically,' namely *disc, disk, monitor*, and *screen*, as illustrated in (3). Another theme concerns STATE OF MIND, which includes noun collocates like *face*, *eye* and *stare*. This theme denotes the meaning of 'showing no emotion, understanding, or interest' as illustrated in (4).

(1) When we were done, our ***pages*** looked ***blank.***

(2) He thought he was staring into ***blank space***, but space was not blank.

(3) Now he switched off his PC and sat gazing at the ***blank monitor***.

(4) … some suggestion of knowledge or warning but his ***face*** was as ***blank*** as ever.

Table 5

Semantic Preference of Noun Collocates of Empty

1. AREA	bed, chair, desk, nest, parking, seat, sky, space, street, tomb
2. RESIDENTIAL AREA	apartment, room
3. CONTAINER	bag, bottle, bowl, box, can, container, cup, glass, plate, shelf, shell, tank
4. CONSUMPTION	beer, calorie, stomach
5. WORTHLESSNESS	promise, suit
6. COMPUTER PROGRAMMING	string

An analysis of the semantic preferences of *empty* shows that there are six major themes as in Table 5. It can be seen that AREA overlaps with that of *blank,* while five new themes emerge from the analysis which include RESIDENTAL AREA, CONTAINER, CONSUMPTION, WORTHLESSNESS and COMPUTER PROGRAMMING. For the first theme, AREA, the noun collocates express 'the area to be used' as illustrated below in (5). The second theme is RESIDENTAL AREA in which the noun collocates specifically indicate 'the place where people live', i.e., *room*, and *apartment* as in (6). The third theme CONTAINER incorporates the majority of nouns indicating 'a hollow object, that can be used for holding something, especially to carry or store it', e.g., *bottle, glass, box*, and *can*, as shown in (7). The fourth theme is CONSUMPTION, which includes *stomach, beer*, and *calorie* as exemplified in (8). The fifth theme, WORTHLESSNESS, contains the noun collocates used with *empty* that express the meaning of something that is 'useless or meaningless', i.e., *promise*, and *suit* as in (9). Finally, the last theme is COMPUTER PROGRAMMING, which includes the word *string* which is used in web coding or computer programming language as shown in (10).

(5) Blackness wrapped around me, and I let myself float through deep, ***empty space***.

(6) The ***apartment*** was ***empty*** and hot, but the hardwood floor felt cool.

(7) … the sake ***bottle*** was ***empty***, and the waitress had brought another

(8) …felt this way at times when I was taking too much or taking on an ***empty stomach***.

(9) Romney was without core or principle, an ***empty suit*** that would say...

(10) So how do you know if the value was null, or truly an ***empty string***?

Table 6

Semantic Preference of Noun Collocates of Vacant

1. AREA	acre, block, chair, desk, land, lot, neighborhood, parcel, parking, seat, shop, space, spot, store, storefront, street, unit, warehouse
2. RESIDENTIAL AREA	apartment, building, home, house, housing, office, property
3. RANK or **JOB**	position, throne, title
4. STATE OF MIND	expression, eye

Table 6 illustrated that semantic preference of noun collocates of *vacant* are categorized into four themes. Only one theme, AREA, is shared with both *blank* and *empty* which indicates that these three words have quite weak near-synonymous status. Most of the nouns are categorized in this theme which include words like *lot, seat, land,* and *space* as in (11). The second theme is RESIDENTIAL AREA, which is also shared with *empty*. The sample sentence from COCA is exemplified in (12). Another theme is RANK or JOB, which includes the noun collocates that express 'social or job position', i.e., *position, title*, and *throne* as shown in (13). The fourth theme, STATE OF MIND, incorporates noun collocates that denote 'showing no expression, awareness, or emotion', namely *eye,* and *expression* as in (14).

(11) View of the address that matches the credit card, and it's just a ***vacant lot*** and a bar

(12) Meanwhile, 24 ***vacant buildings*** and 46 vacant lots are under Ilitch control, …

(13) About 130 people were laid off and other open ***positions*** were left ***vacant***,

(14) … at these moments was frigid and abstract; his ***eyes*** were ***vacant*** in expression

Consequently, by looking at the nouns that *blank*, *empty* and *vacant* usually occur with, more information about the differences in usage between the three adjectives is revealed. In this regard, the distinctions in semantic preference and usage for each synonym based on the collocational differences are supported by previous studies (e.g., Crawford & Csomay, 2016; Gu, 2017; Jirananthiporn, 2018; Li, 2019). Overall, the word *empty* is used in more different themes than *blank* and *vacant*. From the analysis of semantic

preference, these three adjectives share only one theme, i.e., AREA and only one noun collocate, i.e., *space*. This indicates a weak near-synonymous status of the three target words. Furthermore, each adjective is used with the noun collocates depending on the specific contexts. For instance, the notion of job or social position that is 'unoccupied,' as in 'the position was left *vacant'*, is not represented by *blank* and *empty*. However, by looking at each word, it can be seen that certain themes of *vacant* overlap with both *blank,* i.e., STATE OF MIND and *empty*, i.e., RESIDENTIAL AREA. This shows that *vacant* may have a stronger near-synonymous status with both *blank* and *empty*.

Conclusion and Recommendations

This study examined the three synonymous adjectives *blank*, *empty*, and *vacant* based on their degree of formality, and collocational patterns with respect to semantic preference. As the data were gathered from a corpus-based resource, the results of this study demonstrated the frequency of occurrence in different genres and the specific noun collocates of the target words that are not available in dictionaries. The findings illustrated that *empty* occurs with the highest frequency, followed by *blank* and *vacant*, respectively. In terms of their distribution across genres, both *blank* and *empty* most frequently occur in fiction, whereas *vacant* mostly occurs in newspapers. In addition, these synonyms are uncommon in formal English, i.e., academic texts, since this genre appears at the three lowest frequencies. Therefore, these synonyms tend to have quite different occurrences and distributions across genres despite a similar degree of formality. Regarding the noun collocates and semantic preference, these three adjectives have quite weak near-synonymous status as they share only one noun collocate, i.e., *space* and one similar theme of semantic preference, i.e., AREA. Specifically, *blank* collocates with nouns concerning 'space to be written on or filled in' (e.g., *page, check, paper*), while in the context of 'job positions' *vacant* is used (e.g., *position*, *title*). Furthermore, the adjective *empty* occurs in a variety of themes (e.g., residential area, container, worthlessness). It can be concluded that, despite their similarity in meanings, these synonyms co-occur with specific noun collocates in various contexts which cannot be interchangeable (Bailey, 2007; Webb, 2007). For example, *blank* commonly co-occurs with *page* whereas *empty* usually collocates with *bottle.*

Nonetheless, there are some limitations to be considered. First, this study analyzed the data of *blank, empty* and *vacant* solely from COCA which represents American English. It would be interesting to further investigate the use of these target adjectives in other Inner Circle Englishes such as British English using British National Corpus (BNC). In addition, it would be worthwhile to explore and compare the use of these synonyms in varieties

of English, e.g., in ESL and EFL context in order to yield more valid results. Another limitation is that this study examines only three synonymous adjectives. There are other near-synonyms which are also worth exploring using corpus-based data. Additionally, only two criteria were implemented in this study: the degree of formality based on genre distributions and the collocational patterns related to semantic preferences. Other criteria, such as grammatical patterns (Phoocharoensil, 2010), should be used to distinguish the near-synonyms. Lastly, this study determined the collocational strength of near synonymy using frequency and MI score. It is suggested that the collocations derived from a corpus vary depending on the statistical measure used (Phoocharoensil, 2020; Schmitt, 2010). Thus, other statistical methods, such as the z-score, t-score, or log-likelihood tests, should be used in future research to analyze the data.

Pedagogical Implication

Since English synonyms are seen to be a challenge for English language learners, it is crucial that teachers direct their students towards other potentials of language use that might help them become more proficient in English, particularly in terms of word choice. The analysis and results of this study underline the importance of teaching vocabulary in its contexts and using tangible examples to demonstrate how it should be used. In this study, the three target adjectives rarely share similar noun collocates which means they have different semantic preferences. *Blank, empty* and *vacant* may all be used with *space,* but not with other noun collocates. Thus, English Language Teaching (ELT) practitioners can benefit from the study by highlighting the fact that these three synonymous adjectives differ and should be used in appropriate contexts. According to Szudarski (2018), it is important to draw learners' attention to the intrinsic characteristics of each synonym, such as degree of formality, common collocations, and semantic preference. Additionally, the framework from this study can be used by teachers as well as learners to explore additional synonyms that can be used to develop ELT materials for vocabulary lessons, collocation lists as well as data-driven tasks with example sentences that are more applicable to the real-world contexts (Gilquin, 2020; Schmitt, 2010). This will eventually help learners and teachers gain a better understanding of the characteristics of each synonym and be able to use them appropriately.

About the Author

Pallapa Lertcharoenwanich: A lecturer at Business English Department, Buriram Rajabhat University, Thailand. She is currently pursuing a PhD in English Language Teaching at Language Institute, Thammasat University. Her research interests include ELT pedagogy, corpus linguistics and second language acquisition of syntax.

References

Bailey, A. L. (2007). *The language demands of school: Putting academic English to the test*. Yale University Press.

Cai, J. (2012). *Is it "great" enough? A corpus-based study of "great" and its near synonyms* (Publication No. 161807978) [Master's thesis, Ball State University]. Semantic Scholar.

Cambridge dictionary. (Online). https://dictionary.cambridge.org/

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

Clark, E. V. (1992). Conventionality and contrast: Pragmatic principles with lexical consequences. In A. Lehrer & E. F. Kittay, (Eds.), *Frames, fields, and contrasts: New essays in semantic and lexical organization* (pp.171–188). Lawrence Erlbaum.

Collins dictionary. (Online). https://www.collinsdictionary.com

Crawford, W. J., & Csomay, E. (2016). *Doing corpus linguistics*. Routledge.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.

Davies, M. (2020). The corpus of contemporary American English. *English corpora*. https://www.english-corpora.org/coca

DiMarco, C., Hirst, G., & Stede, M. (1993, March 25). *The semantic and stylistic differentiation of synonyms and near-synonyms* [Address]. AAAI Spring Symposium on Building Lexicons for Machine Translation, Stanford, USA. https://www.aaai.org/Papers/Symposia/Spring/1993/SS-93-02/SS93-02-025.pdf

Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics, 28*(2), 105-144. https://doi.org/10.1162/089120102760173625

Firth, J. R. (1957). *A synopsis of linguistic theory 1930-55 in studies in linguistic analysis, philosophical society*. Longman.

Flowerdew, J. (2012). *Corpora and language education*. Palgrave Macmillan.

Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language learning research: Identifying, comparing, and

interpreting the evidence. *Language Learning, 67*(1), 155-179. https://doi.org/10.1111/ lang.12225

Gilquin, G. (2020). Learner corpora. In M. Paquot & S. Gries (Eds), *A practical handbook of corpus linguistics* (pp. 283-303). Springer.

Gu, B. J. (2017). Corpus-based study of two synonyms 'obtain' and 'gain'. *Sino-US English Teaching, 14*(8), 511–522. https://dx.doi.org/10.17265/1539-8072/2017.08.006

Hunston, S. (2007). Using corpus to investigate stance quantitatively and qualitatively. In R. Englebretson (Ed.), *Stance taking in Discourse* (pp. 27-48). John Benjamins.

Jackson, H., & Amvela, E. Z. (2007). *Words, meaning, and vocabulary: Introduction to modern lexicology*. Bloomsbury.

Jirananthiporn, S. (2018). Is this problem giving you trouble? a corpus-based examination of the difference between the nouns 'problem' and 'trouble'. *Thoughts, 2,* 1–25. https://so06.tci-thaijo.org/index.php/thoughts/article/view/163389

Kruawong, T., & Phoocharoensil, S. (2022). A genre and collocational analysis of the near-synonyms teach, educate and instruct: A corpus-based approach. *TEFLIN Journal, 33*(1), 75-97.

Laufer, B. (1990). Ease and difficulty in vocabulary learning: Some teaching implications. *Foreign Language Annals, 23*(2), 147–155. https://doi.org/10.1111/j.1944-9720.1990.tb00355.x

Lewis, M. (2000). *Teaching collocation: Further development in the lexical approach*. Heinle & Heinle.

Li, E. (2019). A corpus-assisted study of synonyms in EFL teaching: Take preserve and conserve as example. *Linguistics and Literature Studies. 7*(2), 39-50. https://doi.org/10.13189/lls.2019. 070201

Lindquist, H. (2009). *Corpus linguistics and the description of English*. Edinburgh University Press.

Liu, D. (2013). Salience and construal in the use of synonymy: A study of two sets of near synonymous nouns. *Cognitive Linguistics, 24*, 67–113. https://doi.org/10.1515/cog-2013-0003

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://doi.org/10.1075/ijcl.17.2.03liu

Longman Dictionary of Contemporary English. (Online). https://www.ldoceonline.com

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

Norris, C. B. (2016). *Academic writing in English*. University of Helsinki.

O'Dell, F., & McCarthy, M. (2008). *English collocations in use: Advanced.* Cambridge University Press.

Oxford advanced learners' dictionary. (Online). https://www.oxfordlearnersdictionaries.com/

Palmer, F. (1981). *Semantics* (2nd ed.). Cambridge University Press.

Partington, A. (2004). Utterly content in each other's company: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics, 9*, 131-156. https://doi.org/10.1075/ijcl.9.1.07par

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.

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. The Southeast Asian Journal of English Language Studies, 26*(3), 1-16. https://doi.org/10.17576/3L-2020-2603-01

Phoocharoensil, S., & Kanokpermpoon, M. (2021). Distinguishing the near-synonyms 'increase' and 'rise': Genre and collocation investigation. *Kasetsart Journal of Social Sciences, 42*(4), 968-975. https://doi.org/10.34044/j.kjss.2021.42.4.34

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

Selmistraitis, L. (2020). Semantic preference, prosody and distribution of synonymous adjectives in COCA. *Journal of Language Studies, 20*(3), 1-18. http://doi.org/10.17576/gema-2020-2003-01

Sinclair, J. (1996). The search for units of meaning. *Textus*, *9*(1), 75-106.

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

Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice.* Routledge.

Vicentini, A. (2003). The economy principle in language notes and observations from early modern English grammars. *Mots Palabras Words, 3*(2003), 37-57. https://www.mediensprache.net/archiv/pubs/3659.pdf

Webb, S. (2007). The effects of synonymy on second-language vocabulary learning. *Reading in a Foreign Language, 19*(2), 120–136.

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics, 27*(1), 103–129. https://doi.org/10.1093/applin/ami045