"A Whole New World… Wait, Is It a *Whole*, *Entire*, or *Total* World?": The Extraction of Collocations for the Three English Synonym Discrimination

PARIWAT IMSA-ARD
Faculty of Education,
Ramkhamhaeng University, Thailand
pariwat.i@rumail.ru.ac.th

SUPAKORN PHOOCHAROENSIL
Language Institute of Thammasat University, Thailand

ABSTRACT

This study investigated three synonymous adjectives, 'whole', 'entire', and 'total', in terms of their collocations and strict sense of meaning. Data were drawn from three English dictionaries: 1) Oxford Advanced Learner's Dictionary (online version), 2) Merriam-Webster's Learner's Dictionary (online version), and 3) Longman Dictionary of Contemporary English (2014), and from the Corpus of Contemporary American English (COCA). The findings from this investigation revealed that the three adjectives share some common meanings, and they can thus be considered near synonyms. They may also be somehow interchangeable with each other in certain contexts, however, in terms of formality, the word 'total' is likely to be used the most in formal contexts. Concerning their common collocations, 'whole' and 'entire' share some noun collocates, some of which are associated with geographic vocabulary; namely America, Europe, and California. However, 'entire' and 'total' share only one strong noun collocate, which is population, while 'whole' and 'total' do not share any typical noun collocates. Significantly, corpus data can provide additional data which does not exist in dictionaries. The findings of this study may serve as supplementary materials for English language teachers to enhance students' English vocabulary learning, especially when it comes to academic writing.

INTRODUCTION

When considering how to communicate effectively in any language, vocabulary is a key and principal component (Gass & Selinker, 2008). Wilkins (1972) demonstrates that "without grammar very little can be conveyed, without vocabulary nothing can be conveyed" (p. 111), indicating that vocabulary is essential to convey meaning in communication. There is an increasing amount of pertinent literature which recognises the importance of vocabulary. It is indisputable that vocabulary plays a pivotal role for learners in acquiring a language. If learners cannot understand the vocabulary from a language they read or hear, then they will not be able to comprehend that language and thus fail to acquire it (Folse, 2004).

In language learning, there are many aspects that L2 learners need to learn such as meanings of words, spoken and written form, collocations, synonyms, connotation, and frequencies (Nation, 2001). In Thailand, English is treated as a foreign language and English language is one of the compulsory courses in Thai education. However, with limited exposure to English, most Thai EFL students still fail in using English in their lives. When comparing the errors Thai EFL students produce, in addition, the errors are concerned with more lexical problems

than syntactic errors. Moreover, many Thai students try to learn new words by studying new words that have exactly the same core meaning in Thai, with a belief that such words can perfectly replace each other and fit well in the sentence. For example, many Thai students are taught by being provided with the words in the same core meaning such as 'provide', 'offer', 'give', 'proffer', 'assign', and 'contribute'. Concerning this, they believe that all these words mean 'to supply' and they interchangeably use any of them in the same context. This can lead to a lexical error in L2. In this regard, using corpora in teaching synonyms can be useful. A corpus is useful in terms of its function in providing lexical information regarding frequencies across genres, collocations, and patterns of grammars.

More specifically to learning vocabulary, learning synonyms is considered one of the most effective ways to expand vocabulary and thus improve writing and speaking. However, one of the major difficulties in English language teaching and learning is that English language vocabulary has innumerable synonyms, and synonymy is therefore considered to be a difficulty in vocabulary learning (Laufer, 1997). To support this claim, Sridhanyarat (2018) demonstrates that synonyms and collocations posed some difficulties for learners' acquisition of second language (L2) vocabulary. Moreover, L2 learners even find learning synonyms confusing and complicated, as they may feel that one word is enough for them. For instance, the adjectives 'whole', 'entire' and 'total' are technically synonyms that correlate with the same meaning of 'all', which is that there is no part left out. However, when examining these adjectives in texts, they seem to differ in terms of their collocational patterns. Moreover, many EFL students and teachers may be confused over the usage of the three adjectives due to their close semantic similarity.

Thus far, no study exploring the synonyms 'whole', 'entire', 'total' appears to have been conducted, and therefore, the similarities and differences of these three synonymous adjectives are worth investigating. Importantly, these three adjectives are listed in the top 2000 most frequent words of the BNC/COCA lists (Nation, n.d.); thus, it is likely that these three words are frequently encountered by language learners due to the fact that the words listed in the top 2000 most frequent words are found in approximately 85-95% in written and spoken English, making them an essential word list for beginning learners seeking to boost their vocabulary growth. Moreover, some scholars (e.g., Ly & Jung, 2015) demonstrate that some limitations of dictionaries could be one of the possible reasons why it is difficult for L2 learners to make a definite distinction among loose-synonyms, causing them to use synonyms in inappropriate contexts. To address these issues, this study focused on the aforementioned three English synonyms of 'whole', 'entire', and 'total', aiming to determine any similarities and differences with regard to their collocational patterns and senses of meaning, and to explore their formality.

LITERATURE REVIEW

ABSOLUTE SYNONYMS VS NEAR SYNONYMS

A synonym refers to words or expressions that have similar or practically the same meaning as another word. In fact, a number of scholars (e.g., Jackson & Amvela, 2000; Cruse, 2000; Elgin, 1992; Taylor, 2002) have attempted to define the term 'synonym'. For instance, Elgin (1992, p. 15) refers to a 'synonym' as "two differing sequences of language that have the same meaning". In addition, Jackson and Amvela (2000) defines 'synonym' as "a relationship of the sameness of meaning that holds between two words" (p. 92). Moreover, Taylor (2002) gives a definition of

'synonym' as words whose semantic similarities are more salient than their differences, and that do not primarily contrast with each other.

There are two distinct categories of synonyms: absolute and near. Absolute synonyms must be identical in all meanings and can be substituted for each other in the same context with the same collocations (Jackson & Amvela, 2000; Lyons, 1995). Near synonyms, also referred to as 'loose synonyms', form the majority of synonyms in the English language. These synonyms share the same or similar definitions, but they cannot be substituted for each other interchangeably in all contexts owing to various factors, such as their scales in meaning and their patterns in use.

In fact, it is indeed difficult to find absolute synonyms which share the same register and style. Cruse (2000) demonstrates that there may not actually be any true absolute synonyms in English. When considering the two synonyms 'big' and 'large', these two English adjectives appear to share the same meaning and register. However, according to the Longman Dictionary of Contemporary English (2014), 'big' refers to something great in size or amount, while 'large' refers to something big in size and quantity. Looking more closely at these two meanings, the slight difference is dealt with by classification, whereby 'big' is used for some degrees and 'large' for some quantities. Therefore, from such a difference, the two mentioned adjectives are not considered absolute synonyms.

DISTINGUIHSING SYNONYMS

There are several criteria used to distinguish synonyms. As two major concerns are investigated in this study, the researchers distinguished synonyms based on two criteria: degree of formality and collocation.

FORMALITY

While synonyms can share similar meanings, they can be different with reference to the degree of formality. Some words may be conventionally used in such formal contexts as academic texts, while others are expected to be used in such informal contexts as spoken language. According to Jackson and Amvela (2000), English slang words, which are widely used in spoken English, are also considered informal words. For instance, the following synonyms are differentiated by their degrees of formality:

Formal Language	Informal Language
simple	cushy
omit	leave out
demonstrate	show
discover	find out

According to pertinent literature, Bailey (2015) points out the features of words in formal contexts in the following criteria:

1) Use of words which originated from Latin and Greek, such as 'attitude' and 'harmony';
2) Formal language, involving the avoidance of personal pronouns to directly refer to the performer of actions, such as 'I' and 'we';
3) Use of precise words, such as 'fifteen years' instead of 'many years';
4) Use of accurate vocabulary, such as the words 'law' and 'rule', which are different.

In addition, Leech and Svartvik (2003) indicate the features of words in informal contexts in the following criteria:

1) Use of words referring to impreciseness, such as 'a lot of' and 'things';
2) Use of idioms or colloquial words, such as 'kids';
3) Use of two-word verbs, such as 'go ahead' or 'bring about';
4) Use of contractions, such as 'can't' instead of 'cannot';
5) Use of personal pronouns, such as 'I think' or 'We believe';
6) Frequent use of wh-question forms, such as 'Why did the pandemic occur?' instead of 'There were five main reasons for the pandemic…'.

COLLOCATIONS

There are several definitions given by various scholars for collocations. According to Benson (1989), "collocations are fixed, arbitrary recurrent combinations of words in which each word basically retains its meaning" (p. 85). Moreover, Nation (2001) and O'Dell and McCarthy (2008) define 'collocations' as the way in which words co-occur or belong to the other words in a natural manner. The adjective 'interested' usually combines with the preposition 'in', and the verb 'communicate' usually co-occurs with the preposition 'with'. As Schmitt (2010) demonstrates, there are two distinct categories of collocations, which are grammatical collocations and lexical collocations.

First, grammatical collocations are comprised of content and grammatical words (specifically prepositions) as follows:

- Noun + preposition: discussion on
- Verb + preposition: abide by
- Adjective + preposition: good at

Second, lexical collocations constitute two equal content words as follows:

- Verb + adverb: sing beautifully
- Verb + noun: pay attention
- Adjective + noun: a significant increase

According to Nation (2001), collocations are significant in language learning. First, stored word sequences are basic language knowledge and usage, and therefore, language knowledge can be considered to be collocational knowledge. Second, to be fluent in English and use appropriate language, collocational knowledge plays an essential role. Also, Kozlowski and Seymour (2003) suggest some significances of collocation learning; for instance, collocations allow EFL/ESL

learners to use words effectively in conveying accurate meaning. Moreover, appropriate collocations help EFL/ESL learners to improve their writing by utilizing more natural language.

PREVIOUS RELATED STUDIES

There have been several research studies scrutinizing adjective synonyms based on corpus-based data drawn from different corpora; for example, the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). Previous literature reveals that common criteria adopted to discriminate synonyms includes meanings and their senses of meanings, collocations, and degree of formality.

Bergdahl (2009) examined the differences of three English synonyms: 'beautiful', 'handsome' and 'good-looking'. Data were drawn from the COCA and five contemporary dictionaries. The findings revealed that the three synonymous adjectives vary in use, despite the fact that they communicate the same core meaning. The adjective 'beautiful' is more associated with women than men, while the adjective 'handsome' is more commonly used with men. Although the adjective 'good-looking' is gender-neutral, it is more commonly used with men according to the COCA.

Another study by Cai (2012) investigated the adjective 'great' along with its six synonyms: 'awesome', 'excellent', 'fabulous', 'fantastic', 'terrific', and 'wonderful', also drawing data from the COCA. The findings revealed that the meaning of 'great' is broader than other synonyms and the adjective 'great' is used more frequently than the other synonyms. The adjectives 'excellent' and 'awesome' are commonly found in magazines, while the adjectives 'great', 'fantastic', 'wonderful', 'fabulous' and 'terrific' appear to be more common in spoken genres.

Hoffmann (2014) investigated six English adjectives, 'kind', 'friendly', 'pleasant', 'nice', 'lovely', and 'gorgeous' based on their semantic and usage differences. In this study, like many other studies, data were drawn from the COCA and three dictionaries. The findings revealed that only some of the adjectives could be used interchangeably in some contexts. The distribution pattern of those collocates were examined to find the degree of formality of the adjectives across eight genres.

In contrast to the previously mentioned studies which did not use inferential statistics in their collocation analysis, Aroonmanakun (2015) conducted a study investigating two English synonyms: 'quick' and 'fast'. Data were drawn from the COCA to explore the similarities and differences between the two synonymous adjectives. The findings indicated that only a few nouns can occur with both adjectives. In terms of meaning, 'quick' indicates the quality of an action while 'fast' provides the sense of the manner of an action.

Later, Petcharat and Phoocharoensil (2017) carried out a study on the three English synonymous adjectives 'appropriate', 'proper', and 'suitable', focusing on their meanings, collocations, degrees of formality, and grammatical patterns. Data were drawn from three dictionaries and the COCA. This study revealed that the three adjectives differed with regard to collocations, degrees of formality, and grammatical patterns, although they shared the same core meaning. The adjective 'appropriate' appears to be more commonly used in formal contexts.

The aforementioned studies highlight how to discriminate synonyms by utilizing different criteria and the sources of both corpora and dictionaries. Corpus studies have become more popular, and the analysis of collocations based on frequencies and mutual information (MI) scores were the motive for this study, to discriminate and differentiate the synonymous adjectives

'whole', 'entire', and 'total', focusing on distribution across eight genres in the COCA. Thus, this study aimed to examine the three English synonyms according to the following research questions:

1. To what extent are the synonymous adjectives 'whole', 'entire', and 'total' distributed across genres?
2. What are the common collocations of the synonymous adjectives 'whole', 'entire', and 'total'?

METHODOLOGY

DATA COLLECTION AND DATA ANALYSIS

In this study, data were primarily drawn from the COCA, which is a very large and genre-balanced corpus of American English. The reason for choosing the COCA is that it is probably one of the most widely used corpora of English in the field of English language teaching, with more than one billion words of text being included over the last two decades. At present, the COCA covers texts from eight different genres; including spoken, TV and Movie subtitles, blogs, webpages, fiction, popular magazines, newspapers, and academic texts. The COCA is believed to be a useful and resourceful corpus for many reasons (Davies, 2010; Hu, 2015). Firstly, the COCA represents American English, and with its enormous size, is comparatively larger than other available American English corpora. Second, the COCA consists of data retrieved from texts equally divided into eight genres. This balance across each genre from year to year can provide data that is relatively different from other corpora. Third, the COCA is a 'monitor' corpus, meaning that the COCA continues to be updated and new texts continue to be added.

A critical component of the COCA's design is that it is fairly equally distributed into eight genres: spoken, fiction, popular magazines, newspapers, academic journals, websites, blogs, and television/movies. To begin with, according to Davies (2020), the spoken genre covers unscripted conversations from over 150 various television and radio shows. Next, the fiction genre contains short stories and plays from literary magazines, children's magazines, and popular periodicals, as well as the first chapters of first edition books published from 1990 to the present, while the popular magazines genre encompasses approximately 100 distinct magazines, with a decent balance of topics such as news, health, home and garden, women, finance, religion, and sports. The newspaper genre includes a diverse range of newspapers from around the United States, with a strong emphasis on local news, opinion, sports, and finance. In addition, the academic journals genre includes approximately 200 peer-reviewed publications and a broad variety of academic disciplines, including education, social sciences, history, humanities, law, medical, philosophy/religion, science/technology, and business. Moreover, the websites genre is classified into academic, argumentative, fiction, informational, instructional, legal, news, personal, promotional, and review web sites, while the blogs genre comprises texts classified as blogs by Google. Finally, the TV/movies genre contains subtitles from OpenSubtitles.org, as well as the TV and Movies corpora, which were added subsequently. The language in these shows or movies is even more colloquial than the data in actual "spoken corpora". The vast set of data for each of these eight genres in COCA demonstrates that it can provide useful information on the frequency of words, phrases, and grammatical structures across genres. To elaborate, this might mean that

they are very informal (e.g., television and movie subtitles or spoken transcripts), more formal (e.g., academic articles), or somewhere in between (e.g., magazines and newspapers).

To address the two research questions, frequencies and distribution across eight different genres of the three synonymous adjectives 'whole', 'entire', and 'total', were first explored in the COCA. For the second research question, in particular, noun collocates frequently co-occurring with the three synonymous adjectives were explored. In this stage, the MI value or score, which regulates "the chance of two words co-occurring in consideration of their frequencies in co-occurring with all the others in the corpus", was used to quantify the collocational strength so as to select typical collocates (Lui & Lei, 2018, p. 6; Phoocharoensil, 2020). However, it should be noted that the MI value has some limitations. According to Cheng (2012) and Szudarski (2018), although some collocations yield high MI scores, they may not be representative due to their low number of occurrences. Therefore, in this study, both frequencies and the MI values were used to measure the strength of collocation. The noun collocates in the top-30 frequency list in the COCA with an MI score ≥ 3 were selected, as this is considered to be a significant value for collocational association (Cheng, 2012; Gablasova et al., 2017).

FINDINGS AND DISCUSSION

To address the two research questions, the similarities and differences in meanings are demonstrated first, followed by the findings concerning the overall frequency of the three synonymous adjectives in the eight genres, the degree of formality in various contexts, and their collocations.

THE SIMILARITIES AND DIFFERENCES IN MEANING BETWEEN 'WHOLE', 'ENTIRE', AND 'TOTAL'

When comparing the definitions of 'whole', 'entire', and 'total' from the Oxford Advanced Learner's Dictionary (online version), the Longman Dictionary of Contemporary English (2014), and the Merriam-Webster Dictionary (online version), data from the three sources show the meanings of the three adjectives to be a full or complete quantity or extent.

TABLE 1. The definition of 'whole', 'entire', and 'total' from three selected dictionaries

words	Oxford Advanced Learner's Dictionary	Longman Dictionary of Contemporary English	Merriam-Webster Dictionary
whole	[only before noun] full, complete e.g., Jenna was my best friend in the <u>whole</u> world.	[only before noun] all of something e.g., I don't believe she's telling us the <u>whole</u> story	having all its proper parts or components e.g., a <u>whole</u> egg
	[only before noun] used to emphasise how large or important something is e.g., We are going to have a <u>whole</u> bunch of people over tomorrow night.	complete and not divided or broken into parts. e.g., a snake swallowing a <u>whole</u> mouse.	constituting an undivided unit e.g., a <u>whole</u> roast suckling pig
	not broken or damaged e.g., Owls usually swallow their prey <u>whole</u>.		seemingly complete or total

			e.g., The <u>whole</u> idea is to help, not hinder.
entire	[only before noun] including everything, everyone or every part e.g., The <u>entire</u> village was destroyed.	[only before noun] used when you want to emphasise that you mean all of a group, period of time, amount, etc. e.g., It was the worst day of my <u>entire</u> life.	having no element or part left out e.g., He was alone the <u>entire</u> day.
			complete in degree e.g., their <u>entire</u> devotion to their family
total	[usually before noun] Being the amount or number after everyone or everything is counted or added together	[usually before noun] complete, or as great as is possible e.g., The sales campaign was a <u>total</u> disaster.	comprising or constituting a whole e.g., the <u>total</u> amount
	used when you are emphasizing something to mean 'to the greatest degree possible' e.g., The room was in <u>total</u> darkness.		absolute, utter e.g., a <u>total</u> failure

According to Table 1, 'whole', 'entire', and 'total' appear to have a similar common core meaning of *being complete*, yet some differences can be identified. The adjective 'whole', generally reveals more detailed meanings than the adjectives 'entire' and 'total' across the three sources. Concerning their usage, the adjectives 'entire' and 'total' are used only before nouns, while the adjective 'whole' does not have to be put before a noun, as in the example '(all) in one piece'.

The findings revealed that the three adjectives also differ in the senses of their meanings. The meaning of 'entire' is likely to be used specifically with a period of time and places, such as 'entire life' or 'entire village'. In addition, 'total' can be used to emphasise absoluteness or the greatest degree of something, such as 'total darkness' and 'total stranger'. Moreover, 'whole' focuses on the completeness of something or emphasises how large or important something is.

Clearly, the findings of this study revealed that the adjectives 'whole', 'entire', and 'total' do not have common exact detailed meanings, and thus cannot be substituted for each other in every context. This finding in terms of meanings is consistent with Cruse (2000) and Jackson and Amvela (2000) indicating that these three synonymous adjectives can be categorised as loose or near synonyms.

FREQUENCY AND DISTRIBUTION

TABLE 2. Overall frequency of three adjectives across the genres

Genre	*whole*		*entire*		*total*		Total frequency
	Frequency	**Per million**	**Frequency**	**Per million**	**Frequency**	**Per million**	
Blog	38,548	299.72	20,829	161.95	16,230	126.19	
Web	36,636	294.85	20,302	163.39	19,151	154.13	
TV	50,083	391.05	13,047	101.87	5,352	41.79	
Spoken	44,563	353.29	13,423	106.42	7,201	57.09	
Fiction	32,944	278.43	12,481	105.48	3,661	30.94	

Magazines	28,429	225.46	17,250	136.81	21,547	170.88	
News	20,405	167.61	11,405	93.68	18,044	148.22	
Academic	19,433	162.22	13,662	114.05	33,625	280.70	
Total	**271,041**		**122,399**		**124,811**		**518,251**

As seen in Table 2, the adjective 'whole' occurs with the greatest frequency (271,041 tokens). In addition, the number of occurrences of the adjective 'whole' (271,041 tokens) is roughly two times higher than that of the adjective 'total' (124,811 tokens), whereas the adjective 'entire' is the least frequent (122,399 tokens) in the corpus data.

TABLE 3. The distribution of the three adjectives across the genres

	whole			*entire*			*total*	
Genre	**Frequency**	**Per million**	**Genre**	**Frequency**	**Per million**	**Genre**	**Frequency**	**Per million**
TV	50,083	391.05	**Web**	20,302	163.39	**Academic**	33,625	280.70
Spoken	44,563	353.29	**Blog**	20,829	161.95	**Magazines**	21,547	170.88
Blog	38,548	299.72	**Magazines**	17,250	136.81	**Web**	19,151	154.13
Web	36,636	294.85	**Academic**	13,662	114.05	**News**	18,044	148.22
Fiction	32,944	278.43	**Spoken**	13,423	106.42	**Blog**	16,230	126.19
Magazines	28,429	225.46	**Fiction**	12,481	105.48	**Spoken**	7,201	57.09
News	20,405	167.61	**TV**	13,047	101.87	**TV**	5,352	41.79
Academic	19,433	162.22	**News**	11,405	93.68	**Fiction**	3,661	30.94
Total	271,041			122,399			124,811	

In Table 3, the adjective 'total' has the highest degree of formality, as its occurrence is the highest and most frequent in academic texts (33,625 tokens with 280.70 per million). The fact that 'total' is mainly associated with formal English is supported by it having the lowest frequency in informal contexts. Specifically, 'total' occurs with the lowest frequency in fiction (3,661 tokens with 30.94 per million), TV and movie subtitles (5,352 tokens with 41.79 per million) and spoken (7,201 tokens with 57.09 per million), all of which are considered to be informal English genres. Likewise, the frequency of 'entire' is very low in fiction (12,481 tokens with 105.48 per million) and TV and movie subtitles (13,047 tokens with 101.87 per million), and 'whole' appears to have the highest degree of informality as its occurrence is the highest in TV and movie subtitles (50,083 tokens with 391.05 per million) and spoken (44,563 tokens with 353.29 per million), all of which are also considered to be informal English genres. Regarding formality, 'entire' seems to be more formal than 'whole', as indicated by its frequency in the academic genre, while 'total' is apparently the most formal.

Clearly, the findings corroborate with pertinent literature suggesting the adoption of genre differences as a criterion to distinguish synonyms (Jackson & Amvela, 2000; Phoocharoensil, 2020). Thus, the three synonymous adjectives 'whole', 'entire' and 'total' are commonly used in different genres, signaling that they are near-synonyms and cannot be used interchangeably in every context.

TABLE 4. Nouns collocates of 'whole', 'entire', and 'total' in COCA

Rank	whole			entire			total		
	Noun collocate	Frequency	MI Score	Noun collocate	Frequency	MI Score	Noun collocate	Frequency	MI Score
1	thing	28,074	3.28	population	1,748	3.49	number	5,521	3.96
2	lot	11,246	3.46	career	1,587	3.41	percent	3,509	3.68
3	bunch	3,098	4.85	planet	878	3.31	fat	3,190	6.51
4	grain	1828	5.39	universe	816	3.74	cost	2,593	4.34
5	wheat	1,122	5.43	us	596	4.85	score	2,111	5.04
6	mess	986	3.73	length	570	3.52	population	2,038	4.29
7	lotta	332	7.10	spectrum	282	3.69	min	1,864	7.12
8	slew	274	5.54	continent	265	3.97	amount	1,631	4.03
9	cloth	258	3.03	ecosystem	258	3.99	yard	1,283	4.31
10	shebang	253	8.52	cast	253	3.88	return	1,042	4.19
11	ordeal	248	4.68	fleet	239	3.93	stranger	908	5.21
12	clove	225	4.03	genome	169	4.07	revenue	882	4.46
13	genome	216	3.51	America	162	5.90	loss	865	3.27
14	America	189	5.21	premise	153	3.08	income	846	3.67
15	hog	174	3.96	semester	126	3.27	percentage	839	4.54
16	fiasco	152	4.66	California	100	7.05	carbohydrate	836	6.23
17	ballgame	119	5.52	clan	92	3.21	sample	783	3.79
18	gamut	113	5.16	hemisphere	88	3.42	sales	681	3.62
19	enchilada	103	5.77	apparatus	87	3.70	lack	641	3.40
20	raft	100	3.20	cosmos	71	3.70	share	637	3.62
21	peppercorn	97	5.77	ordeal	70	3.59	debt	625	3.46
22	debacle	96	3.69	wardrobe	66	3.24	asset	615	4.31
23	Europe	79	6.51	Africa	56	6.86	cholesterol	613	5.55
24	charade	68	4.31	Europe	56	6.93	offense	607	4.39
25	California	65	5.52	Sunday	55	6.16	calorie	535	4.12
26	edifice	62	4.50	watershed	54	3.56	spending	473	3.90
27	sequencing	60	3.14	Mexico	51	6.63	expenditure	434	5.74
28	legume	52	4.28	repertoire	50	3.36	failure	427	3.01
29	Sunday	49	5.08	edifice	50	5.10	variance	410	5.58
30	neighbourhood	44	3.39	Chicago	48	6.44	output	365	4.88

According to Table 4, it is evident that 'whole' and 'entire' seem to be more synonymous, as these two words share certain common collocates. The six noun collocates that 'whole' and 'entire' share are *America, Europe, California, edifice, genome* and *Sunday*, where the first three noun collocates are related to geographic vocabulary. However, the only noun collocate that 'entire' and 'total' share is *population*, while 'whole' and 'total' do not share any noun collocates from the top-30 frequency list. In fact, there could be more noun collocates shared by these three synonyms, but due to the exclusion of those with an MI score lower than 3, some may be absent from Table 4. Therefore, it is also significant to acknowledge that the shared noun collocates are limited to only those presented in Table 4.

Analyzing the semantic preference of the synonymous words, the noun collocates can be organised into groups based on their similarities in meaning. Hunston (2002) refers to semantic preference as "the frequent co-occurrence of a lexical item with items expressing a particular evaluative meaning" (p. 266). That is to say, semantic preference is the connection between a set of semantically related words and word forms (Begagić, 2013). To analyse semantic preference, Ly and Jung (2015) demonstrated that information about collocations can help determine the association ranges of words searched, as well as the semantic relationships among their collocates.

TABLE 5. Semantic preference of noun collocates of 'whole'

1.	**PLACE**	shebang, America, Europe, California, edifice, neighbourhood
2.	**PLANT**	grain, wheat, clove, peppercorn, legume
3.	**FOOD**	enchilada
4.	**STATE**	fiasco, debacle
5.	**ACTIVITY**	slew, ordeal, ballgame, charade, raft
6.	**MISCELLANEOUS**	thing, lot, bunch, mess, lotta, cloth, genome, hog, gamut, sequencing, Sunday

All the noun collocates were grouped based on their semantic preference. In Table 5, the noun collocates of 'whole' are categorised into six themes, namely PLACE, PLANT, FOOD, STATE, ACTIVITY and MISCELLANEOUS. The majority of its noun collocates belongs to PLACE and MISCELLANEOUS. Looking at the table 5, it is clear that the noun collocates of 'whole' appear to be associated with places, plants and food, as illustrated below:

- The **whole** *room* seems noisy when she comes in.
- I buy **whole** *bean coffee* and grind it to put in that.
- I ate the **whole** *croissant* and then I was eyeing the next one too.

TABLE 6. Semantic preference of noun collocates of 'entire'

1.	**SCIENCE**	planet, universe, spectrum, ecosystem, genome, hemisphere, cosmos, apparatus
2.	**PLACE**	continent, America, California, Africa, Europe, Mexico, edifice, Chicago, U.S., premise
3.	**PEOPLE**	population, fleet, clan
4.	**EVENT**	ordeal, watershed
5.	**MISCELLANEOUS**	cast, semester, length, career, wardrobe, Sunday, repertoire

In Table 6, the noun collocates of 'entire' were categorised into five themes; namely SCIENCE, PLACE, PEOPLE, EVENT, and MISCELLANEOUS. The majority of its noun collocates belong to SCIENCE, followed by PLACE. The salient majority of noun collocates of 'entire' somehow appear to accord with the noun collocates of 'whole'. Evidently, the noun collocates of 'entire' are mostly related to scientific and geographic nouns as illustrated below:

- Metagenomics is the study of the genomic features of **entire** *bacterial communities*.
- It looked like the **entire** *city* would be filled with snarled traffic.
- Only a fraction of all public land in the **entire** *country*, including Colorado.

TABLE 7. Semantic preference of noun collocates of 'total'

1.	**BUSINESS**	cost, amount, revenue, income, sales, debt, asset, spending, expenditure, return, loss, share
2.	**MATHEMATICS**	number, percent, score, min, percentage, variance, output
3.	**SCIENCE**	fat, carbohydrate, cholesterol, calorie
4.	**PEOPLE**	population, stranger, sample
5.	**MISCELLANEOUS**	lack, offense, failure

In Table 7, the noun collocates of 'total' were categorised into five themes, namely BUSINESS, MATHEMATICS, SCIENCE, PEOPLE, and MISCELLANEOUS. The majority of its noun collocates belong to BUSINESS, followed by MATHEMATICS. The noun collocates of 'total' are mostly associated with numbers as follows:

- You can dramatically reduce the **total** *amount* you pay on your loan.
- …as percentages of the **total** Asian *population*…
- …a document that outlines the county's **total** *spending* annually…

DEGREE OF FORMALITY

The degree of formality of the texts was analysed based on the key features suggested by Bailey (2015) and Leech and Svartvik (2003) as discussed in the literature review. The findings revealed that the COCA highlights the degree of formality, which is demonstrated in Table 8.

TABLE 8. The percentages of concordance lines of the three adjectives in formal and informal contexts

| Synonyms | Total concordance lines of each synonym: 200 | | | |
| | Formal | | Informal | |
	No. of lines	Percent	No. of lines	Percent
whole	76	38.00%	124	62.00%
entire	108	54.00%	92	46.00%
total	129	64.50%	71	35.50%

After analyzing the degree of formality of each concordance line (see Figure 1) based on the criteria for distinguishing formal and informal language of Bailey (2015) and Leech and Svartvik (2003) as discussed in the literature review, Table 8 demonstrates the distribution of concordance lines for each synonym in both formal and informal contexts. It is evident that 'whole', 'entire', and 'total' occur in both formal and informal contexts. Concerning the formal contexts, 'total' scores the highest (129 lines), followed by 'entire' (108 lines) and 'whole' (76 lines). Regarding informal contexts, 'whole' scores the highest (124 lines), while 'entire' and 'total' constitute 92 and 71 lines respectively. In addition, the findings illustrate that the three words in this study are not similar with regard to their degree of formality. To elaborate, 'total' appears to be used more frequently in formal contexts than the other two adjectives. These findings seem to corroborate the results from the distribution of three adjective synonyms across the genres presented in Table 3.

FIGURE 1. The example of the KWIC concordance line of 'whole'

However, 'entire' and 'total' are not as significantly different in degree of formality as 'entire' and 'whole', and 'total' and 'whole'. To elaborate, the number of concordance lines for each synonym makes up about 50% of the total lines found in both formal and informal contexts, in accordance with the study by Petcharat and Phoocharoensil (2017). This indicates that although 'entire' and 'total' are different concerning their degrees of formality, the differences are not so clearly apparent as that of 'whole' and 'entire', and 'whole' and 'total'.

The findings of this present study suggest that corpus-based study can provide more insightful information than can be obtained solely from a dictionary, especially in terms of usage in various contexts (Aroonmanakun, 2015; Hunston, 2002). In this study, the findings reveal that the three synonymous adjectives 'whole', 'entire' and 'total' are considered 'near-synonyms' and are in agreement with the study by Jarunwaraphan and Mallikamas (2020) and Phoocharoensil (2020) in which they concluded that near-synonyms may be used differently in terms of collocation and semantic preferences although they share similar meanings. This is further supported by other linguists (e.g., Stubbs, 2001; Taylor, 2003), demonstrating that true synonyms are rare, and thus all synonyms are truly near-synonyms. Significantly, other studies (e.g., Liu, 2010; Uba, 2015) investigating adjective synonymy revealed synonymous adjectives to often not be entirely identical in meaning, and hence not completely interchangeable.

CONCLUSION

This study revealed that three English synonyms 'whole', 'entire', and 'total' demonstrate some similarities concerning meaning. The data were drawn from the Corpus of Contemporary American English (COCA) and three dictionaries. Despite the fact that they share some core

meanings, the three English synonyms differ with regard to collocations and their degree of formality. Concerning the formality, the adjective 'total' appeared to be used the most in formal contexts. Regarding the collocations, the adjectives 'whole' and 'entire' seem to be more synonymous, because these two adjectives share certain common collocates. The six noun collocates that 'whole' and 'entire' share are *America, Europe, California, edifice, genome* and *Sunday*. Based on the collocations, the first three noun collocates are associated with geographic vocabulary. The only noun collocate that 'entire' and 'total' share is *population*, while 'whole' and 'total' do not share any noun collocates in the top-30 frequency list. In this study, it is proffered that data from corpora can provide additional details which cannot be found in a dictionary. Moreover, this study demonstrates that corpora can be beneficial, as they serve as a useful language teaching material for language teachers, and it is important that teachers should teach their students to be mindful of such distinctions between synonymous words and how to use them appropriately.

However, it should be noted that there are some limitations in this study. First, the conclusions from this investigation are mainly based on interpretations of evidence from corpus concordances lines, and the English language may not truly be reflected by a sole corpus (Hunston, 2002). Second, there are only three synonyms examined in this investigation. Therefore, further investigations should include other words in the set of synonyms; for example, 'complete' or 'full'. Finally, the number of the noun collocates are limited to those in the top-30 frequency lists. An inclusion of other noun collocates beyond the top-30 frequency list may establish a clearer and more comprehensive picture of collocational patterns of investigated synonyms.

PEDAGOGICAL IMPLICATIONS

It is important for English language teachers to apply corpus knowledge concerning synonyms to their teaching, such as the findings and the procedures from this investigation. Several studies revealed that language used in coursebooks is frequently based on intuitions rather than actual evidence of use from corpus data (O'Keeffe et al., 2007). For instance, Gilmore (2004) found that textbook dialogues differ considerably from their naturally occurring equivalents. As the findings suggested, some words may seem to provide similar meanings, but they may not be interchangeable with one another in certain contexts. To support this, Szudarski (2018) demonstrated that a lack of awareness of the characteristics of synonyms could lead to producing unnatural language. As mentioned in the title of this research study, 'whole' and 'entire' may be substituted for the word 'world' perfectly, but 'total' would not be used in reality. Consequently, the meaning of words obtained exclusively from dictionaries may not be sufficient to suggest their use in natural and real contexts.

ACKNOWLEDGEMENT

REFERENCES

Aroonmanakun, V. (2015). Quick or fast: A corpus based study of English synonyms. *LEARN Journal: Language Education and Acquisition Research Network, 8*(1), 53-62.

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

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

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

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

Cai, J. (2012). *Is it "great" enough? A corpus-based study of "great" and its near synonyms* [Master's Thesis]. Ball State University, Indiana, USA.

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

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

Davies, M. (2010). The corpus of contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing, 25*(4), 447-464. https://doi.org/10.1093/llc/fqq018

Davies, M. (2020). *The COCA corpus (new version released March 2020)*. https://www.english-corpora.org/coca/help/coca2020_overview.pdf

Elgin, S. H. (1992). *What is linguistics?*. Prentice Hall.

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

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

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

Gilmore, A. (2004). A comparison of textbook and authentic interactions. *ELT Journal, 58*(4), 363-374.

Hoffmann, K. (2014). A corpus-based analysis of the near synonyms 'Nice', 'Kind', 'Lovely', 'Friendly', 'Gorgeous', and 'Pleasant. In *122228 BA Paper–Usage-based approach to linguistic investigations* (pp. 1-29). Universität Wien, Austria.

Hu, H. C. (2015). A semantic prosody analysis of three adjective synonymous pairs in COCA. *Journal of Language and Linguistic Studies, 11*(2), 117-131. https://dergipark.org.tr/en/pub/jlls/issue/36118/405581

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

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

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

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

Kreidler, C. W. (1998). *Introducing English semantics*. Taylor & Francis Group.

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

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.

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

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

Liu, D., & Lei, L. (2018). Using corpora to teach vocabulary. *The TESOL Encyclopedia of English Language Teaching*, 1-8.

Longman (2014). *Longman dictionary of contemporary English* (6th ed). Pearson Education Limited.

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

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

Merriam-Webster dictionary. https://www.merriam-webster.com/

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

Nation, P. (n.d.). *Vocabulary lists*. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-lists.

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

O'Keeffe, A. & McCarthy, M. (2010). *The Routledge handbook of corpus linguistics*. Routledge.

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

Oxford advanced American dictionary. https://www.oxfordlearnersdictionaries.com/definition/american_english/

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

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

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

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

Sinclair, J. (2004). *Trust the text*. Routledge.

Sridhanyarat, K. (2018). Thai learners' acquisition of L2 collocations: An interlanguage perspective. *GEMA Online Journal of Language Studies, 18*(1), 1-21. https://doi.org/10.17576/gema-2018-1801-01

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

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

Taylor, J. R. (2002). Near synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Science, 25*, 263-284.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Edward Arnold.

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