

Primary, Main, and Major: Learning the Synonyms through Corpus Data

Supakorn Phoocharoensil

supakorn.p@litu.tu.ac.th

Language Institute of Thammasat University, Thailand

ABSTRACT

English is widely known as a language containing a number of near-synonyms, i.e. words with similar meaning, and therefore English learners are often confronted with difficulty in the use of near-synonyms in different contexts. This corpus-informed study aims to differentiate the synonyms *primary*, *main*, and *major*, focusing on their distribution across genres and collocation usage. The three target synonyms were selected as *main* and *major* are among the first 1,000 words in spoken and written English (Longman Dictionary of Contemporary English, 2014), while *primary* appears in Coxhead's (2000) Academic Word List (AWL). The data from the Corpus of Contemporary American English (COCA) demonstrated that the three synonymous adjectives most frequently occur in academic texts, with *primary* being more common in newspapers and magazines. Through a collocation analysis based on frequency and MI score, *concern* strongly collocates with all three adjectives, while some noun collocates are frequently combined with only certain pairs of synonyms, e.g. *primary/main focus*, *a main/major theme*, and *a primary/major factor*. More interestingly, some noun collocates are attached to specific semantic themes, with *primary* being exclusively associated with health and election, *main* with place, food, or literature, and *major* with sports or business. As to the implications for ELT, teachers are encouraged to develop synonym lessons based on typical collocates derived from authentic corpus-based English, e.g. COCA.

Keywords: English synonym; synonymous adjective; collocation; genre; COCA

INTRODUCTION

Vocabulary, often viewed as "...the fuel of language, without which nothing meaningful can be understood or communicated" (Gardner, 2013, p. 2), plays a crucial role in English language education, and there is a clear link between vocabulary knowledge and learners' development of the four main skills (Nation, 2013). A primary challenge facing English learners in L2 vocabulary acquisition is synonymy, i.e. the lexical relationship in which two or more linguistic forms have the same meaning (Carter, 2012). As illustrated by Phoocharoensil (2020a), although two words share some semantic similarities (e.g. *consequence* vs. *outcome*), the substitution of one by the other may affect the naturalness in L2 English production since particular combinations (e.g. *a dire consequence*) apparently form strong collocational patterns that are more widely accepted than others (e.g. **a dire outcome*). Learners whose L2 experience or exposure is limited can find it difficult to determine which synonyms are most appropriate in specific contexts.

Many studies have focused on the discrimination of English near-synonyms in a variety of linguistic aspects, including degree of formality (e.g., Jirananthiporn, 2018; Phoocharoensil, 2020b, 2021a, 2021b), connotations (e.g., Partington, 1998; Phoocharoensil, 2020a; Stubbs, 1995), collocations (e.g., Crawford & Csomay, 2016; Jirananthiporn, 2018; Phoocharoensil, 2020a,

2020b, 2021a, 2021b), semantic prosody (e.g., Nelson, 2006; Partington, 2004; Phoocharoensil, 2021a, 2021b; Selmistraitis, 2020), and colligations (e.g., Phoocharoensil, 2021a, 2021b).

The main focus of this study is to investigate three synonymous adjectives, *primary*, *main*, and *major*, with respect to their distribution across genres and collocations. A comprehensive analysis of the three target synonyms based on the data from an enormous, reliable corpus representing American English, i.e. COCA, can, to a large extent, enhance the generalizability of the findings as regards the proper usage of each individual synonym. The cognitive meanings and collocational information of the three adjectives from a learner's dictionary, i.e. Longman Dictionary of Contemporary English (2014), were examined to confirm the synonymy of the target words, and then the actual occurrences of these synonyms in various COCA text types were taken into consideration. Next, extraction of noun collocates frequently co-occurring with the adjective synonyms was performed to allow for a more detailed comparison between the semantic themes to which the collocates were assigned.

The current study selects three high-frequency adjective synonyms, i.e. *primary*, *main*, and *major*, as confirmed by Longman Dictionary of Contemporary English (2014), with *main* and *major* being among the first 1,000 words in spoken and written English and *primary* being included in the first 2,000 written English words in Coxhead's (2000) Academic Word List (AWL).

LITERATURE REVIEW

This section introduces the concept of synonymy and the ways to distinguish synonyms using different criteria.

SYNONYMY

Synonymy is a very important concept in lexicology as well as language teaching. According to Carter (2012), synonymy refers to a "symmetrical sense relation in which more than one linguistic form can be said to have the same conceptual or propositional meaning" (p. 34). Examples of synonyms are *house*, *home*, *abode*, and *domicile*, all of which refer to 'the place where someone lives' (Carter, 2012). Two major categories of synonyms exist: perfect synonyms and near-synonyms. 'Perfect' or 'absolute synonyms', on one hand, are defined as words whose senses of meaning are identical and thus can be substituted in all possible contexts of use with no effect on the original meaning, style, or connotation (Cruse, 1986). In reality, such perfect or strict synonyms in this sense are exceedingly rare or even non-existent, for redundancy will arise when a language appears to have two words whose meanings are truly identical (Jackson & Amvela, 2007).

In contrast to perfect synonyms, 'loose synonyms', also commonly known as 'near-synonyms', are defined as words that are semantically close and for which the meaning overlap can be identified or differentiated in context (Phoocharoensil, 2020a, 2020b). As Jackson and Amvela (2007) note, near-synonyms are not always substitutable in every context. A very clear example of near-synonyms is *prize* and *award*, both referring to 'something given to someone to reward them for something they have done'. They are possible alternatives to each other in (1), wherein both are "similar enough to be judged synonymous" (Murphy, 2009, p. 137); meanwhile, in (2), *award*, rather than *prize*, is the best choice to fit this specific context. It is noteworthy that when the concept 'synonym' is generally referred to in ELT, it concerns the varying degrees of near-synonyms, as opposed to absolute synonyms.

- (1) Jan won the **prize/award** for the best drawing
(Murphy, 2009, p. 137)
- (2) The plaintiff received a hefty **award** (**≠prize**) in the lawsuit.
(Murphy, 2009, p. 137)

The ability to distinguish near-synonyms in the same set is of paramount importance for ELT teachers and learners. Although words can be very similar in denotative meaning, their usage often differs, which often generates confusion among English users, including teachers not speaking L1 English (Phoocharoensil, 2020a). It is the context that helps determine the right synonyms to be selected on a particular occasion (Carter, 2012; Murphy, 2009). Linguists rely on certain criteria to draw subtle distinctions between near-synonyms, some of which are worth discussing in the following subsection.

SYNONYM DISTINGUISHING CRITERIA

Lexicologists apply certain criteria to illustrate how near-synonyms are used differently depending on the context (Jackson & Amvela, 2007). First, it is possible to make a distinction between synonyms by looking at connotations. That is, in referring to a particular thing, two words might share some core meaning, but “they may have divergent associative or emotive meanings” (Jackson & Amvela, 2007, p. 96). The adjectives *little* and *small* in (3) and (4), respectively, are good examples of synonyms differing in connotations. While their denotative meanings overlap to a great degree, i.e. ‘not large in size or amount’ (LDOCE, 2014), *little* sounds more generous than *small*, probably because “it[*little*] has a more emotive, positive quality related to its use as an endearment” (Murphy, 2009, p. 154).

- (3) The employees received a **little** Christmas bonus.
- (4) The employees received a **small** Christmas bonus.

The second criterion relates to the style or formality of the context in which a word occurs. In a pair of synonyms, one word may be more associated with a high level of formality, whereas the other could be more commonly used in informal contexts. For instance, even though *warning* and *caveat* are synonymous, *warning* is more frequently found in a less formal context than the latter (LDOCE, 2014). Some corpus-based studies have considered degree of formality in synonym differentiation. In Jirananthiporn (2018), the noun *problem* is apparently attached to text types in COCA that represent a greater degree of formality, whereas its synonym *trouble* seems to be more widely used in less formal genres. Also, based on COCA data, Jarunwaraphan and Mallikamas (2020) discovered that the synonymous nouns *chance* and *opportunity* have different distributions across genres in that the highest frequency of *chance* was observed in the spoken genre, whereas *opportunity* appears more often in academic texts. Phoocharoensil (2020b) revealed that the noun *error*, which occurs with the highest frequency in academic texts in COCA, is obviously characteristic of formal English, while its synonyms *fault* and *mistake* were found to occur most frequently in TV/movie subtitles, i.e. one of the most informal genres in this corpus. In a similar vein, it was reported in Phoocharoensil (2021b) that the verb synonyms *foresee* and *predict* appear to have a marked difference in formality, with *predict* occurring with the highest frequency in academic texts against *foresee* being the most common in webpages, a genre that is relatively less formal than academic language.

Another critically important criterion for differentiating near-synonyms is collocations. According to Baker, Hardie, and McEnery (2006), the term ‘collocation’ in corpus linguistics is defined as “the phenomenon surrounding the fact that certain words are more likely to occur in combination with other words in certain contexts” (p. 36). Collocations are normally measured by statistical means, e.g. MI scores (or mutual information), the Z-score, log-likelihood, etc. (Flowerdew, 2012; Saito, 2020). Murphy (2009) maintains that “words tend to pattern with limited ranges of other words” (p. 156), which means that despite sharing the same conceptual meaning, synonyms seem to differ in terms of the collocational patterns in which they usually occur. By way of illustration, although both *rancid* and *addled* can refer to food that has gone bad, the two synonyms exclusively co-select different nouns to follow. It is much more natural in English to use the collocations *rancid bacon* and *addled eggs* rather than **rancid eggs* and **addled bacon*. Such collocational restrictions prevent substitution of synonyms in every context, supporting Thornbury’s (2002) claim that “Even the slightest adjustments to the collocation – by substituting one of its components for a near-synonym...turns the text into non-standard English” (p. 7).

Typical collocations of a word can be revealed only through consultation of a large native-speaker corpus. As shown in Murphy’s (1998) study based on information from language corpora, collocations help measure the degree of similarities between near-synonyms. The study indicated that the adjectives *big* and *enormous* are close synonyms as they both share several contexts in which substitution of either is allowed. By contrast, considering the collocates with which they co-occur, *large* and *enormous*, despite their meaning overlap, were found in none of the same environments. It is clear that “no two words can be considered perfect synonyms as corpus data reveal important differences in the phraseological patterns...” (Szudarski, 2018, p. 43).

Many studies use corpus-based approaches in collocational analysis to account for dissimilarities in the usage of near-synonyms. Synonymous nouns have been the primary focus of several subsequent studies (e.g. Jarunwaraphan & Mallikamas, 2020; Phoocharoensil, 2020a, 2020b, 2021a). Having analyzed the common verb and adjective collocates of *opportunity* and *chance* in the classic version of COCA (i.e. comprising five genres), Jarunwaraphan and Mallikamas (2020) point out that the verb collocates of *opportunity* tend to convey positive connotations, e.g. *create*, *expand*, *extend*, *promote*, and *provide*, while *chance* often co-selects verbs that largely involve negative connotations, e.g. *damage*, *destroy*, *eliminate*, *jeopardize*, and *reduce*. With regard to the adjective collocates, *opportunity* seems to be mainly combined with adjectives linked to positive rather than negative contexts, e.g. *adequate*, *convenient*, *favorable*, *important*, and *tremendous*, whereas *chance* strongly collocates with adjectives conveying mixed connotations, e.g. *ample*, *high*, and *numerous* (positive) vs. *limited*, *low*, and *slim* (negative), due to the word’s polysemous properties.

In more recent studies, in-depth analysis of typical collocations has been facilitated by the advent of the most updated version of COCA, consisting of eight genres as of March 2020 (Davies, 2020). Phoocharoensil (2020a) investigated the common adjectives and verbs collocating with the nouns *consequence*, *result*, and *outcome*. The findings demonstrated that *consequence* noticeably expresses negative senses through its adjective collocates, e.g. *adverse*, *catastrophic*, *dire*, *disastrous*, *fatal*, *grave*, *harmful*, *severe*, and *tragic*, and verb collocates, e.g. *befall*, *evade*, *face*, and *suffer*. In comparison with *consequence*, *result* seems to co-occur with research-oriented verbs, e.g. *corroborate*, *generalize*, *interpret*, and *replicate*, and adjectives, e.g. *descriptive*, *empirical*, *experimental*, *mixed*, *qualitative*, and *quantitative*, while *outcome* combines with verb and adjective collocates with a wide variety of semantic properties. Phoocharoensil (2020b) conducted another study on noun synonym discrimination involving three target words: *error*, *mistake*, and

fault. It was shown in COCA that *error* and *mistake* share far more adjective collocates (e.g. *fatal*, *glaring*, *grave*, and *grievous*) and verb collocates (e.g. *correct* and *rectify*) than *fault*, meaning that these two nouns are closer near-synonyms, compared with *fault*, which presents a more limited range of collocates.

Another study devoted to synonymous verb differentiation by Phoocharoensil (2021a) focused on *persist* and *persevere*, both sharing the cognitive meaning of ‘to continue doing something in a difficult situation’. A thorough investigation indicated that the verbs are associated with different contexts, as clearly shown by the dissimilar sets of noun collocates. That is, *persist* tends to collocate with nouns denoting negative/adversative meanings, e.g. *symptom*, *rumor*, *gap*, *myth*, *drought*, *inequality*, *tension*, *poverty*, and *racism*, whereas *persevere* frequently co-exists with 1) Christian-oriented vocabulary, e.g. *God*, *saints*, *prayer*, *disbelieved*, *guard against evil*, *in the midst of dryness*, *spiritual aridity*, and *a Catholic school*, and 2) phraseological units reflecting difficulties and determination, e.g. *hike*, *setback*, *marathon*, *obstacles*, *difficulties*, *disappointment*, *hard work*, *withstanding the burden*, *battles you must face*, and *have faced much criticism*. A set of synonymous verbs, i.e. *teach*, *educate*, and *instruct*, was examined, with an emphasis on object noun collocates, by Kruawong and Phoocharoensil (2022). The objects of *teach* are often school or university subjects, e.g. *biology*, *economics*, *English*, *mathematics*, *physics*, while those of *educate* involve the interaction of society and social work, e.g. *citizen*, *communities*, *disabilities*, *immigrants*, *people*, *workers*, *youth(s)*. The object noun collocates frequently combined with *instruct* are usually associated with legal English, e.g. *agents*, *attorneys*, *judges*, *jury/jurors*.

The current study adopts two major criteria, namely the degree of formality as illustrated in the distribution across genres, as well as collocational patterns, to distinguish between the synonymous adjectives *primary*, *main*, and *major*. The two research questions which the study aims to address are as follows:

1. How are the synonyms *primary*, *main*, and *major* distributed across different genres?
2. What are the common noun collocates with which the synonyms *primary*, *main*, and *major* frequently co-occur?

METHOD

DATA COLLECTION & DATA ANALYSIS

For this study, the Corpus of Contemporary American English (COCA) was consulted. COCA is one of the most popular and widely used corpora representing American English. As a huge genre-balanced corpus, COCA contains more than one billion words of text, with around 25 million words having been added on an annual basis from 1990 to 2019 in eight genres: 1) five conventional genres, i.e. spoken, fiction, popular magazines, newspapers, and academic texts; and 2) three new genres, i.e. TV/movie subtitles, blogs, and other web pages, as of March 2020 (Davies, 2020). Alongside other sizable corpora constituting the ‘mega-corpora’ of BYU (Brigham Young University), COCA has been extensively used in research and English Language Teaching (ELT) for a number of reasons (Friginal, 2018). Researchers, firstly, can explore the frequency of search words across eight well-balanced genres in quest of their common collocates, and even compare the different lexical uses between genres. Second, as a monitor corpus, new data are added annually to COCA, constantly enlarging its size; this makes it a reliable source of contemporary native-speaker English that can effectively inform ELT practitioners. Another primary reason why COCA is widely used in ELT concerns the fact that it significantly promotes

students' autonomous and inductive learning, known as data-driven learning (DDL), through the presentation of authentic English data (Yamtui & Phoocharoensil, 2019).

Due to all the aforementioned reasons, COCA was selected to distinguish the near-synonyms *primary*, *main* and *major*. With *Word*, i.e. one of the new functions available in the latest version of COCA (Davies, 2020), researchers are allowed to access the basic information of the word being searched, including its distribution across genres, definitions, related topics, collocates, etc., as long as the search words are among the top 60,000 words in COCA. As noted by Ma and Mei (2021), COCA users can “understand the meaning nuances and use patterns for the keyword and its related word” (p. 184), thereby enabling them to successfully and systematically differentiate near-synonyms based on the typical collocates and the genre(s) with which each synonym is associated.

The present study aims to provide answers to the two above research questions. COCA was first consulted for frequencies and distribution across genres of the target synonyms, i.e. *primary*, *main* and *major*, in each of the eight genres. As for the second research question, the noun collocates frequently accompanying the three noun synonyms were explored. The typical collocates were extracted based on the collocational strength as rated by the Mutual Information (MI) score in conjunction with the frequency. Although high MI scores seem to indicate a strong collocational association, collocation selection based on only the MI score is inappropriate since many collocations with very high MI scores could occur with a low total frequency in a corpus (Gablasova, Brezina & McEnery, 2017). Put differently, collocation research relying on MI scores alone may focus more on rare combinations, as the MI value tends to promote such words in ranking the collocates (Barnbrook, Mason, & Krishnamurthy, 2013). Students are very unlikely to encounter such rare, peculiar collocations in daily life. Teachers, hence, do not need to incorporate these into their vocabulary lessons (Szudarski, 2018). For this study, the collocational strength was measured based on both the frequency of noun collocates as the principal collocation-extraction criterion, along with the MI scores, to ensure that only recurrent, common collocates would be extracted. Accordingly, the top-20 high-frequency noun collocates shown in COCA whose MI score was ≥ 3 , i.e. the significance value for collocational strength (Cheng, 2012), were listed.

In the following step, the extracted common noun collocates of *primary*, *main* and *major* were categorized according to the semantic preference. In particular, collocates sharing meaning similarities were assigned to the same category. The researcher then explored the noun collocates that frequently co-occur with all or some of the target synonyms, as this can be evidence of some semantic overlap of the selected synonyms; furthermore, the nouns that seem to collocate specifically with certain synonyms were highlighted to clearly show how each synonym co-selects its particular collocates.

RESULTS AND DISCUSSION

Table 1 shows the shared meanings of the target synonyms *primary*, *main* and *major*, sentence examples, and the common noun collocates of these three synonyms obtained from Longman Dictionary of Contemporary English or LDOCE (2014), in which all three words are considered near-synonyms of one another:

TABLE 1. Meanings and usage of the synonymous adjectives *primary*, *main*, and *major* in LDOCE

	<i>primary</i>	<i>main</i>	<i>major</i>
meaning	most important	larger or more important than all other things, ideas etc. of the same kind	very large or important, when compared to other things or people of a similar kind
sentence example	Our <i>primary</i> concern is to provide the refugees with food and health care. (LDOCE, 2014, p.1435)	The <i>main</i> reason for living in Spain is the weather. (LDOCE, 2014, p.1102)	Britain played a <i>major</i> role in the negotiations. (LDOCE, 2014, p.1103)
common collocations	<i>primary</i> purpose/aim/objective	<i>main</i> reason/problem/concern/ point/ aim/bedroom	<i>major</i> role/part/factor

In the following table, the entire frequency of the target synonyms is displayed according to the genres in which they occur.

TABLE 2. Distribution of the Synonyms *Primary*, *Main*, and *Major* across Genres according to Frequency

Genre	<i>primary</i>		Genre	<i>main</i>		Genre	<i>major</i>	
	Frequency	Per million		Frequency	Per million		Frequency	Per million
academic texts	25,002	208.71	academic texts	19,997	166.93	academic texts	37,314	311.49
webpages	11,020	88.69	webpages	17,959	144.53	newspapers	36,611	300.73
newspapers	8,552	70.25	newspapers	16,102	132.26	magazines	31,065	246.37
blogs	8,874	69.00	blogs	16,381	127.37	webpages	25,754	207.27
magazines	8,252	65.44	magazines	14,956	118.61	spoken blogs	24,625	195.23
spoken	8,183	64.87	fiction	10,652	90.03	TV and movies subtitles	24,129	187.61
TV and movies subtitles	1,541	12.03	spoken	9,214	73.05	TV and movies subtitles	10,307	80.48
fiction	1,364	11.53	TV and movies subtitles	5,364	41.88	fiction	6,980	58.99
Total	72,788		Total	110,652		Total	196,785	

Table 2 shows the total frequency of the three adjective synonyms. Overall, *major* has the highest raw frequency (196,785 tokens), followed by *main* (110,652 tokens) and *primary* (72,788 tokens), respectively. It is clear that these three synonyms are more common in written English, as can be seen in the top-5 genres, all of which represent written language. The three words are widespread in academic texts, i.e. the genre in which they occur with the greatest frequency, with *major* being the most frequent (311.49 per million), followed by *primary* (208.71 per million) and *main* (166.93 per million), respectively. Their high frequency in academic texts indicates that the three synonyms have a high level of formality. The two synonyms *primary* and *main* are similar in their distribution across genres, sharing the same ranks in the following genres: academic texts, webpages, newspapers, blogs, and magazines, respectively. In comparison with *primary* and *main*, *major* is more frequent in newspapers (300.73 per million) and magazines (246.37 per million), respectively. The prevalence of the target synonymous adjectives in written English also correlates

with their low frequency in spoken English, TV/movie subtitles, and fiction, i.e. genres closely associated with informal, colloquial language.

TABLE 3. Noun Collocates of *Primary*, *Main*, and *Major* in COCA

Rank	Noun collocater	<i>primary</i>		Noun collocater	<i>main</i>		Noun collocater	<i>major</i>	
		Frequency	MI Value		Frequency	MI Value		Frequency	MI Value
1	care	4371	5.15	character	4,925	5.18	league	4852	5.38
2	source	2304	4.88	reason	4,405	3.83	change	3471	3.02
3	goal	1599	4.56	effect	2,627	3.75	role	2767	3.31
4	reason	1441	3.25	street	2,427	3.40	factor	2204	3.57
5	focus	1106	5.20	road	1,674	3.33	concern	1625	3.22
6	concern	1103	4.34	course	1,477	3.61	corporation	1046	3.95
7	purpose	1090	4.41	event	1,460	3.05	component	870	3.64
8	physician	973	5.86	source	1,456	3.35	theme	846	3.48
9	election	831	3.45	goal	1,336	3.43	newspaper	825	3.00
10	responsibility	735	4.19	menu	1,316	5.69	championship	758	3.86
11	objective	649	5.55	concern	1,294	3.65	shift	740	3.66
12	function	604	4.19	focus	1,139	4.35	contributor	734	5.13
13	factor	508	3.08	entrance	891	5.40	label	621	3.85
14	voter	501	3.54	purpose	889	3.22	airline	613	3.83
15	mission	471	3.48	theme	692	3.95	contribution	576	3.12
16	target	447	3.57	objective	647	4.68	depression	558	3.29
17	outcome	425	3.85	dish	610	4.09	obstacle	557	4.44
18	prevention	424	5.23	attraction	606	5.41	manufacturer	478	3.52
19	provider	371	4.71	gate	524	3.86	initiative	453	3.02
20	caregiver	361	6.71	ingredient	460	4.29	breakthrough	447	4.91

The nouns strongly collocating with the adjectives *primary*, *main*, and *major* as determined by the MI value of ≥ 3 at the minimum, i.e. the level of statistical significance for collocational association, are listed in Table 3. As near-synonyms whose core meanings are close, the three target adjectives obviously have some noun collocates in common. This seems to confirm that the selected adjectives are synonyms of one another (Phoocharoensil, 2020a, 2020b). As reported in Table 3, many nouns frequently co-occur with *primary* and *main*, i.e. *concern*, *focus*, *goal*, *objective*, *purpose*, and *source*, while those commonly used with both *primary* and *major* are fewer, i.e. *concern* and *factor*. The noun collocates that *main* and *major* share are *concern* and *theme*. The noun *concern* is apparently the only collocate often combined with all the three adjectives. However, a careful interpretation of the corpus-based information in Table 3 is required when dealing with common collocations. To illustrate, as only the top-20 nouns were targeted in the extraction of collocations, which is a limitation of this study, there are likely to be other nouns that can also co-occur with the three target adjectives but are not shown in any of the three lists above because their frequency or MI score in COCA is low. For instance, the collocation *major goal* occurs 326 times in COCA but was not selected because the frequency is not among the top 20; meanwhile, the number of occurrences of *primary color* is high in COCA (590 tokens), but its MI score of 2.90 resulted in its exclusion from the collocation list. It should also be noted here that though many basic nouns, e.g. *things*, can definitely co-occur with the target words, as in *primary things*, *main things*, and *major things*, such weak collocations were left unlisted because they are less problematic for learners and thus not considered pedagogically useful (Hill, 2000).

TABLE 4. Semantic Preference of Noun Collocates of *Primary*

1. AIM & FOCUS	<i>focus, goal, objective, purpose, target</i>
2. DUTY	<i>concern, function, mission, responsibility</i>
3. HEALTH	<i>care, caregiver, physician, provider</i>
4. CAUSE	<i>factor, reason, source</i>
5. ELECTION	<i>election, voter</i>
6. MISCELLANEOUS	<i>outcome, prevention</i>

Upon a thorough analysis of the semantic preference of the noun collocates of primary, five principal themes emerged, as revealed in Table 4. The first theme AIM & FOCUS includes *focus, goal, objective, purpose, and target*, as exemplified in (5), and the second DUTY contains noun collocates that are similar in meaning, i.e. *concern, function, mission, and responsibility*, as shown in (6). It is interesting to highlight two themes that are characteristic of the adjective primary, namely HEALTH (including *care, caregiver, physician, and provide*, as in (7) and CAUSE, including *factor, reason, and source*, as in (8). The fifth theme, containing *election and voter*, as exemplified in (9), is ELECTION. Two nouns that do not seem to fit any one of the themes, *outcome and prevention*, have been assigned to MISCELLANEOUS. Nonetheless, it is worth noting that new themes can arise when more possible noun collocates are included in future studies. This way, the collocates currently belonging to MISCELLANEOUS could be assigned to one of the emerging themes.

if additional noun collocates are extracted in future studies, these two nouns could be included under new themes that may arise.

- (5) The Columbia Land Conservancy's primary focus is on facilitating leases rather than sales.
- (6) Believe it or not, my primary concern is making the world a cleaner place.
- (7) ... President Reagan developed Alzheimer's, and the first lady became his primary caregiver, until his death in 2004.
- (8) It's absolutely crystal clear that the primary factor driving the collapse of the nuclear renaissance is economics.
- (9) In most states, territories, and the District of Columbia, candidates for the U.S. House of Representatives who are members of major political parties are nominated in a primary election.

TABLE 5. Semantic Preference of Noun Collocates of *Main*

1. AIM & FOCUS	attraction, focus, goal, objective, purpose, theme
2. PLACE	entrance, gate, road, street
3. FOOD-RELATED MATTERS	course, dish, ingredient, menu
4. CAUSE	reason, source
5. EFFECT	effect
6. DUTY	concern
7. LITERATURE	character

Table 5 presents seven themes to which the noun collocates of *main* have been assigned. Three of the themes related to *main* co-exist with those characterizing *primary*, i.e. AIM &

FOCUS, including *attraction, focus, goal, objective, purpose, and theme*, as in (10), CAUSE, including *reason and source*, as in (11), and DUTY, including *concern*, as in (12).

(10) The tours introduce artists to the community and may result in sales, but that's not the **main goal** of the event...

(11) Presenting the evidence to substantiate this argument is not the **main reason** the book was written.

(12) Fiscal conservatism and reduction of debt should be our **main concern**.

While the theme EFFECT is also one of the key themes of *major*, to be discussed below, three more themes are specifically associated with *main*, namely PLACE, FOOD-RELATED MATTERS, and LITERATURE. To be more precise, the noun collocates connected with PLACE are *entrance, gate, road, and street*, as exemplified in (13), and those under FOOD-RELATED MATTERS encompass *course, dish, ingredient, and menu*, as in (14). Importantly, LITERATURE contains only one member character, the combination *main character*, as can be seen in (15), and it is the most frequent in COCA.

(13) The campground was about one mile south on the **main road**.

(14) Try eating a healthy snack prior to the **main course**, such as a piece of fruit.

(15) There's a scene where the **main character** is sitting in class with his cell phone out, playing around with it.

TABLE 6. Semantic Preference of Noun Collocates of *Major*

1. BUSINESS	airline, corporation, manufacturer, newspaper
2. EFFECT	breakthrough, change, contribution, shift
3. DUTY	concern, initiative, role
4. CAUSE	contributor, factor
5. SPORTS	championship, league
6. PROBLEM	depression, obstacle
7. AIM & FOCUS	theme
8. MISCELLANEOUS	component, label

As illustrated in Table 6, three themes concerned with *primary* and *main* were also discovered for *major*, namely DUTY, CAUSE, and AIM & FOCUS. The nouns typically collocating with *major*, as opposed to the other two adjectives, are *initiative* and *role*, as in (16), under DUTY, and *contributor*, as in (17), under CAUSE, while the noun *theme*, as in (18), under AIM & FOCUS, is also a common collocate of *main*. Furthermore, compared to *effect*, the single noun collocate of *main* under EFFECT, those of *major* comprise four nouns, i.e. *breakthrough, change, contribution, and shift*, as shown in (19), rather than *effect*.

(16) There's a good chance price plays a **major role** in the decision not to adopt or upgrade service.

(17) As part of the water cycle, groundwater is a **major contributor** to flow in many streams and rivers and has a strong influence on river and wetland habitats for plants and animals.

(18) The book's **major theme** centers on what it really means to be a disciple of Christ.

(19) This work represents a **major breakthrough** in the field of nanobiotechnology as it demonstrates the ability to leverage recent advances in the field of DNA origami pioneered by researchers around the world...

The themes that deserve special attention are BUSINESS, in which *airline*, *corporation*, *manufacturer*, and *newspaper* are members, as in (20), SPORTS, including *championship* and *league*, as in (21), and PROBLEM, with *depression* and *obstacle* as its members, as in (22), none of which were found to be linked to the other two synonyms. This finding tremendously facilitates discrimination between the three target near-synonyms.

(20) As a CEO of a ***major corporation***, I had to stand in front of thousands of shareowners and take questions.

(21) It was a given that these two proud and fierce competitors would play the match as though it were a ***major championship***, and they did.

(22) Ineffective communication between parents and teachers can be a ***major obstacle*** when trying to solve problems with students, but fortunately it can be improved.

As illustrated in the findings of this study, the three synonyms, despite the overlap in their core meanings, most commonly occur in formal, written English, with academic texts being the genre in which they are found with the highest frequency. The adjective *major* seems to be far more prevalent in newspapers and magazines than *primary* and *main*. The lowest frequency of *main* was observed in TV/movie subtitles, whereas fiction is the text type where *primary* and *major* are the least common. The fact that some synonyms are more common in specific genres accords with several past studies (e.g. Jackson & Amvela, 2007; Jirananthiporn, 2018; Phoocharoensil, 2020a, 2020b, 2021a, 2021b).

In addition to the frequency across genres in COCA, this study yields a sharper distinction between the collocational patterns of the three adjective synonyms. With the typical collocates being extracted and then classified into various themes, both the similarities and differences in the synonyms' collocational behavior can be observed in a clearer fashion. In terms of similarities, the frequent noun collocates of *primary*, *main*, and *major* share the same themes of DUTY, CAUSE, and AIM & FOCUS. The COCA data, nevertheless, indicate that the individual collocates under the same theme partially differ. For example, while *primary* and *main* typically co-occur with the nouns *focus*, *goal*, *objective*, *purpose*, the noun *theme* seems to form a strong collocation with *main* and *major*, rather than *primary*. Close scrutiny revealed the differences in the characteristics of the noun collocates of each synonymous adjective. For instance, some nouns appear to be particularly combined with *primary* in referring to 'health', e.g. *physician*, and *provider*, and 'election', e.g. *election* and *voter*, whereas some nouns exclusively co-exist with *main*, as opposed to *primary* and *major*, referring to 'place', e.g. *entrance*, *road*, and *street*, 'food', e.g. *course*, *dish*, and *ingredient*, and 'literature', e.g. *character*. As for *major*, there are distinctive noun collocates widely used in the contexts of 'business', e.g. *airline* and *corporation*, and 'sports', e.g. *championship* and *league*. The existence of such variation in the typical collocations closely related to each synonymous adjective clarifies the differences in the usage between the three synonyms, which is nearly impossible to explain without sufficient authentic data from corpora (e.g. Crawford & Csomay, 2016). The three synonyms are made distinguishable given that typical, specific collocates are explicitly displayed (e.g. Jirananthiporn, 2018; Kruawong & Phoocharoensil, 2022; Phoocharoensil, 2020a, 2020b, 2021a, 2021b). In other words, the findings are consistent with the claim that "there are important differences between the words that tend to be perceived as synonyms and access to corpora provides us with an opportunity to explore such differences" (Szudarski, 2018, p. 43).

CONCLUSION & PEDAGOGICAL IMPLICATIONS

As the frequency across genres clearly demonstrated, all three target adjectives are prevalent in academic language, i.e. the text type with a very high level of formality, with *major* being more common than the others in newspapers and magazines. The relative infrequency of the three synonyms in TV/movie subtitles and fiction seems to confirm that the three words are associated with formal English. Furthermore, classifying the adjectives by their collocations enabled the researcher to detect the similar patterns they share, and more importantly, the different collocational behavior. This means that although the same nouns, e.g. *concern*, can often be used with all three target synonyms or at least two, e.g. *theme* as a typical collocate of *main* and *major*, particular semantic categories are attached to each adjective. For example, some collocates of *major* are linked to sports or business, while some nouns frequently appearing with *primary* are concerned with health or elections. Regarding *main*, there are some collocating nouns that relate to place, food, or literature. Without the considerable amount of collocational data from a vast corpus like COCA, it is unlikely that such instructive results regarding the semantic themes of the noun collocates of the target adjectives would have been brought to light (Phoocharoensil, 2020a, 2020b).

ELT practitioners may view the findings of this study as useful in several ways. First, based on the abovementioned results, teachers can share with students a clear, convincing explanation of the subtle distinctions between *primary*, *main*, and *major* with respect to genre distribution and collocations (Phoocharoensil, 2021a). It is advisable that teachers, relying upon the corpus data of native-speaker English, emphasize to students the fact that few or no absolute synonyms exist in actuality, and almost all synonyms in English are regarded as loose synonyms, which can never be interchangeable in all contexts of use (Jackson & Amvela, 2007). Second, teachers can address other groups of synonyms using a similar corpus-based method, and incorporation of common collocations derived from authentic language in corpora will substantially enhance their ELT material development, facilitating students' synonym acquisition through context-based natural English (Friginal, 2018; Szudarski, 2018).

The findings of this study, however, have to be considered in light of some limitations. First, this study extracted the top-20 noun collocates based on frequency and MI score. Unlike some previous studies in which the possible typical collocates were limited in number (e.g. Phoocharoensil, 2020a), the current study focused on grouping 20 collocates according to their semantic preferences, resulting in some main themes under which these collocates were grouped. The selection of additional noun collocates from COCA would likely generate more themes, potentially yielding interesting empirical results. It is anticipated that other noun collocates, aside from concern, can co-occur with the three synonymous adjectives. Another significant limitation concerns the criteria used in distinguishing the synonyms. This study relied on two criteria, namely the degree of formality as indicated by the distribution across genres and the common collocations. Future researchers can explore and compare recurrent lexical bundles made up of such adjective synonyms, e.g. *our main objective is to* or *one of the main goals*, in different disciplines. Employing knowledge of lexical bundles in which one synonym but not others can also be a key component should enable researchers to effectively discriminate synonyms.

REFERENCES

- Baker, P., Hardie, A. & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Barnbrook, G, Mason, O. & Krishnamurthy, R. (2013). *Collocation. Applications and Implications*. London: Palgrave Macmillan.
- Carter, R. (2012). *Vocabulary. Applied Linguistic Perspective*. London: Routledge.
- Cheng, W. (2012). *Exploring Corpus Linguistics. Language in Action*. London: Routledge.
- Coxhead, A. (2000). A New Academic Word list. *TESOL Quarterly*. 34(2), 213-238.
- Crawford, W. J. & Csomay, E. (2016). *Doing Corpus Linguistics*. London: Routledge.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge: Cambridge University Press.
- Davies, M. (2020, June 25). *The New Corpus of Contemporary American English (COCA 2020)*. Language Institute Thammasat University (LITU) Webinar, Bangkok, Thailand. <https://www.youtube.com/watch?v=zW9I9C9rPD4>
- Flowerdew, L. (2012). *Corpora and Language Education*. London: Palgrave Macmillan.
- Friginal, E. (2018). *Corpus Linguistics for English Teachers*. London: Routledge.
- 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
- Gardner, D. (2013). *Exploring Vocabulary in Action*. London: Routledge.
- Hill, J. (2000). Revisiting Priorities: From Grammatical Failure to Collocational Success. In M. Lewis (Ed.). *Teaching Collocation: Further Development in the Lexical Approach* (pp. 47-69). London: Commercial Colour Press Plc.
- Jackson, H. & Amvela, E. (2007). *Words, Meaning, and Vocabulary: An Introduction to Modern English Lexicology*. London: Cassel.
- Jarunwaraphan, B. & Mallikamas, P. (2020). A Corpus-Based Study of English Synonyms: Chance and Opportunity. *rEFLections*, 27(2), 218-245. <https://so05.tci-thaijo.org/index.php/reflections/article/view/248710>
- Jirananthiporn, S. (2018). Is This Problem Giving You Trouble? A Corpus-Based Examination of the Differences between the Nouns ‘Problem’ and ‘Trouble’. *Thoughts 2018*. 2, 1–25.
- 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.
- Longman Dictionary of Contemporary English* (2014). London: Pearson Education.
- Ma, Q. & Mei, F. (2021). Review of Corpus Tools for Vocabulary Teaching and Learning. *Journal of China Computer-Assisted Language Learning*, 1(1), 177–190. <https://doi.org/10.1515/jccall-2021-2008>
- Murphy, M. L. (1998, Oct. 9-11). *What Size Adjectives Tell us about Lexical Organization* [Conference session]. The Linguistic Association of the Southwest Conference, Tempe, Arizona, United States.
- Murphy, M. L. (2009). *Semantic Relations and the Lexicon*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486494>
- Nation, I. S. P. (2013). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge: Cambridge University Press.
- Nelson, M. (2006). Semantic Association in Business English: A Corpus-Based Analysis. *English for Specific Purposes*, 25, 217–34. <https://doi.org/10.1016/j.esp.2005.02.008>

- Partington, A. (1998). *Patterns and Meanings. Using Corpora of English Language Research and Teaching*. Amsterdam: John Benjamins.
- Phoocharoensil, S. (2020a). 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.
- Phoocharoensil, S. (2020b). Collocational Patterns of the Near-Synonyms *Error, Fault, and Mistake*. *The International Journal of Communication and Linguistic Studies*, 19(1), 1–17.
- Phoocharoensil, S. (2021a). Semantic Prosody and Collocation: A Corpus Study of the Near-Synonyms *Persist* and *Persevere*. *Eurasian Journal of Applied Linguistics*, 7(1), 240–258.
- Phoocharoensil, S. (2021b). Multiword Units and Synonymy: Interface between Collocations, Colligations, and Semantic Prosody. *GEMA Online® Journal of Language Studies*, 21(2), 28–45.
- Saito, K. (2020). Multi- or Single-Word Units? The Role of Collocation Use in Comprehensible and Contextually Appropriate Second Language Speech. *Language Learning*, 70(2), 548–588.
- Selmistraitis, L. (2020). Semantic Preference, Prosody and Distribution of Synonymous Adjectives in COCA. *GEMA Online® Journal of Language Studies*, 20(3), 1–18. <http://dx.doi.org/10.17576/gema-2020-2003-01>
- Szudarski, P. (2018). *Corpus Linguistics for Vocabulary: A Guide for Research*. London: Routledge.
- Thornbury, S. (2002). *How to Teach Vocabulary*. Harlow: Longman.
- Yaemtui, W. & Phoocharoensil, S. (2019). Effectiveness of Data-Driven Learning on Enhancing High-Proficiency and Low-Proficiency Thai EFL Undergraduate Students' Collocational Knowledge. *Asian EFL Journal*, 23(3.2), 290–314.

ABOUT THE AUTHOR

Supakorn Phoocharoensil is Associate Professor of English from Language Institute of Thammasat University, Thailand. He is currently teaching various undergraduate courses in English for Specific Purposes, and MA. and Ph.D. courses in English Language Teaching and Applied Linguistics. His areas of research interest include Second Language Acquisition, Corpus Linguistics, and English Collocations.