Type, Kind, or Sort: Exploring Near Synonyms through Corpus Analysis

Panupan Panthong[1]/ Naviya Chutopama[2]/ Thitaporn Putklang[3]

1Business English Program, Faculty of Humanities and Social Sciences, Buriram Rajabhat University

2,3English Program, Faculty of Humanities and Social Sciences, Buriram Rajabhat University

1E-mail: panupan.pt@bru.ac.th

2E-mail: naviya.jc@bru.ac.th

3E-mail: thitaporn.pk@bru.ac.th

Abstract

One of the key problems in learning vocabulary among second language learners is near synonyms. Previous studies explored different sets of near synonyms to contribute evidence-based knowledge for L2 learners and classroom implementations. This study aims to investigate the differences among three synonyms including Type, Kind and Sort and study the noun and adjective collocations in relation to semantic preferences. Using the Corpus of Contemporary American English, this study examines the frequency of these words across various genres and their noun and adjective collocations, selected based on the significant MI score level of ≥ 3 and categorized by theme. The findings found that these synonyms, despite some similarities, cannot be used interchangeably as their usage varies across genres. The word 'Type' is predominantly linked with medical contexts while 'Kind' and 'Sort' often appear in discussions about groupings and classifications. The study emphasizes the importance of incorporating these insights into teaching materials and classroom strategies to enhance the vocabulary acquisition process for L2 learners, ensuring a more precise and comprehensive grasp of language usage.

Keywords

Statements of the problems

In the research of second language acquisition as a branch of applied linguistics, the research area of vocabulary was marginalized and recognized as a 'Cinderella subject' (Daller et al., 2007; Milton, 2009; Milton & Hopwood, 2022). Nonetheless, this research area goes beyond the recognition and becomes more pompous, shining the light on teaching and research for second and foreign language learners. According to the classic quote by Wilkins (1972), the absence of vocabulary can hinder the learners from understanding everything. Therefore, vocabulary plays an indispensable role for learners to become more fluent in the target language (Celen & Yalçin, 2021; Laufer, 1990; Yaacob et al., 2019). In the classroom, it is important to present the learners to use different words but the same meaning (synonyms) because the use of such various words could enhance learners' lexical repertoire (Ayhan Yavuz, 2014; Kamali Khalavi & Zeraatpishe, 2023).

However, one of the key problems in learning vocabulary in a second or foreign language is near synonym. According to Petcharat and Phoocharoensil (2017), synonym falls into two categories: strict and loose synonyms. The former type refers to words that share the same meaning and concept and can be used interchangeably, but the number of this type of synonyms is relatively rare (Chung, 2011; Taylor, 2003). On the other hand, loose synonyms are sometimes recognized as near synonym (Petcharat & Phoocharoensil, 2020; Phoocharoensil, 2013, 2020b, 2021a, 2021b; Phoocharoensil & Kanokpermpoon, 2021; Sridhanyarat & Phoo- charoensil, 2023). They seem challenging to second language learners. That is, the learners may not be able to use the words in the contexts because there are some differences among these words for learners to understand despite using the dictionary (Lertcharoenwanich, 2023; Sridhanyarat & Phoocharoensil, 2023). Phoocharoensil (2020a) gave a set of examples of near synonyms which are 'consequence', 'result', and 'outcome'. These words share the same meanings of concept, but they are used differently with collocations.

Previous studies have researched near synonyms in different sets of words. For exam- ple, Lertcharoenwanich and Phoocharoensil (2022)studied two near synonyms 'persist' and 'preserve' in COCA corpus, and he also analyzed the semantic prosody by collocation of these two words. Similarly, other works also undertook the similar approaches with other sets of near synonyms such as 'error', 'fault' and 'mistake' (Phoocharoensil, 2020), persist' and 'preserve' (Phoocharoensil, 2021), 'increase' and 'rise' (Phoocharoensil & Kanokpermpoon, 2021), 'edu- cate', 'teach' and 'instruct' (Kruawong & Phoocharoensil, 2022), 'assess', 'evaluate', and 'mea- sure' (Sridhanyarat & Phoocharoensil, 2023), 'big' and 'large' (Aimjirakul 2015), 'problems' and

'trouble' (Jirananthiporn 2018), 'important', 'significant' and 'crucial' (Chaokongjakra 2023), and 'blank', 'vacant' and 'empty' (Lertcharoenwanich, 2023). These studies contributed a great deal of knowledge of near synonyms in different genres and also the collocation to reflect the semitic preferences in the contexts. However, relatively rare research was conducted on type, kind and sort.

'Type', 'kind' and 'sort' are considered as near synonyms. As illustrated in Table 1, the meanings of these three words in two different online dictionaries show that they encompass the concept of categorization and classification. The definition of each word is likely to blur the line between the sensitivity of meaning in the use of different contexts. According to Phoocharoensil (2021), the second L2 learners are stymied by near synonym because they are not aware of such words. Moreover, the selection of these three words is based on the reason that they are in the General Service List (GSL) (see West, 1953). As experts in vocabulary research (Nation, 2001; Schmitt & Schmitt, 2014) suggested, teaching words in this list as a group of high frequency words could enhance the learners' comprehension. Apart from that, Laufer (1997) also recognised this list as the foundation to access the learners to the vocabulary knowledge. Therefore, this study aims to study this set of near synonyms to benefit L2 learners of how different these words are and to inform results for language teachers to implement in the classroom.

Table 1: The Definitions of Type, Kind and Soft in Two Online Dictionaries

Words	Oxford Advanced Learner's Dictionary	Cambridge Dictionary
type	a class or group of people or things that share particular qualities or features and are part of a larger group; a kind or sort	a particular group of people or things that share similar characteristics and form a smaller division of a larger s
kind	a group of people or things that are the same in some way; a particular variety or type	a group with similar characteristics, or a particular type
sort	a group or type of people or things that are similar in a particular way	a group of things that are of the same type or that share similar qualities

Research Objectives

1. To investigate the differences among three synonyms e.g., type, kind and sort

2. To study the noun and adjective collocations in relation to semantic preferences of type, kind and sort.

Scope of the study

Although there are many different criteria to distinguish near synonyms (e.g., dialect, formality, connotation, collocational pattern and semantic preference), this present study adopted Phoocharoensil's (2020a) framework including a degree of formality, collocational pattern, and semantic preferences. These three criteria are likely to contribute the results of differences among these words.

Methodology
Sample

According to the research objectives in the aforementioned section, the sample of the study includes three near synonyms including 'type', 'kind' and 'sort'. These three sample words were adopted from the Corpus of Contemporary American English or COCA which includes more than 1.1 billion words collected between 1990 and 2019 and consists of eight genres including academics, website, blog, magazine, spoken, news, TV and fiction (see english-corpora.org/coca/)

The rationale for choosing this COCA is because of the following reasons. First, it is considered as a big corpus containing many different text types or genres, so it could be beneficial for this present study to explore the differences among genres and identify the formality of words. Second, the data of this corpus could be used to improve the vocabulary skills of learners. According to Fang, Ma and Yan (2021), learners improved their vocabulary skills after the training session of the COCA use. Moreover, Koçak, (2020) found that learners see the COCA corpus as a benefit tool for the references to learn a language. Finally, the selection of COCA allows a present study to discuss the results of previous studies.

Data collection

Because this present study adopted the data from the ready-to-use corpus website of COCA, the process of data collection is related to browsing target words and collecting them as an electronic file. As a result, the data collection introduces the following stages.

The first stage is to access data on the COCA official website. The researchers visited the website and clicked on the Browse section to type each target word. After that, the data of each word that appeared upon the search were collected based on their eight genres including academics, website, blog, magazine, spoken, news, TV and fiction. Next, the frequencies of each word are arranged from the most frequent to the least frequent occurrence in the table to show the frequency in the different genres in the Excel file.

The next step is to search for the collocation of each word. The researchers clicked on Collocates in order to collect the data of collocations to further investigate the semantic preferences. The result of this search presents the frequency and MI scores of a list of collocations, but not all of them are included in this study. Only the frequency of the collocation is not likely to indicate the level of occurrence. It should be noted that the MI scores which can reflect the collocation strength as the representative (Cheng, 2012). When it comes to an analysis of collocations of words, corpus linguistics experts (Biber, 2006, 2009; Gablasova et al., 2017) suggested that Mutual Information or MI scores should be used as a criterion. For example, Cheng (2012) assigned that the significant level of MI score to analyze the collocation is ≥ 3. For this reason, the researcher excluded unrelated collocation by clicking the Advance Option to select the minimum MI scores of each word.

     The last step is to reach the concordance lines of collocations in each word to show examples in relation to the semantic preference. To do so, the researchers clicked on the icon after the list of collocations to see the words in context and then concordance lines in KWIC functions appeared. After that, the researchers read line by line to examine the semantics preferences of collocations.

Data Analysis

     The data analysis in this present study adopted Phoocharoensil's (2020) framework, introducing two stages based on the research objectives. The first stage is related to the degree of formality while another stage is related to the thematic analysis of semantics preferences based on the collocations that appeared in concordance lines.

The first objective is to investigate the differences among three synonyms e.g., type, kind and sort. As previously mentioned in the first stage of data collocation, both the overall frequency and frequency of different gernes were examined by the data from COCA. The data presented based on their frequency could indicate the degree of formality.

     The second stage is related to the exploration of the noun and adjective collocation in relation to semantic preference. It should be noted that only collocations with MI score value (≥ 3) are included in this study. Normally, previous studies (Lertcharoenwanich & Phoocharoensil, 2022; Phoocharoensil, 2020a) which investigated nouns of near synonyms analyze adjective and verb collocations. However, our preliminary investigation found that verb collocations do not meet at MI score value (≥ 3) and there is only 1 adverb collocation of each word. Therefore, this study excludes an analysis of verb and adverb collocations because they

are relatively rare in the corpus. Next, the thematic analysis is conducted to categorize, and ground based on semantics preference. As a result, this examination could differentiate the differences among these three words.

Results & Discussion

After the process of data collection and analysis, this present study presents the results based on the research objectives. The first subsection reports the distribution among type, kind and sort in different genres in COCA and then the second subsection informs the result of noun and adjective collocation in relation to their semantic preference.

Differences among three synonyms

In response to the first objective, the result revealed that the frequency of these three words vary across genre, so it could be interpreted that they are not able to be used interchangeably. Overall, the highest frequency of all genres is 'kind', followed by 'type' and 'sort'. Table 2 shows that the frequency of type (27,911 times) in the academic genre is more frequent than that of kind (19,984 times) and that of sort (7,202 times. On the website, 'kind' with 41,155 frequencies is the most common, followed by 'type' and 'sort' with 22,222 and 16,682, respectively. In a blog genre, kins occurs 47,170 times while 'sort' occurs 20,155 times and 'type' occurs 18,944. Also, the number of occurrences of the word 'kind' outnumbers that of 'type' and 'sort' in four genres i.e., magazines, news, and fiction and TV. However, both 'kind' (117,536) and 'sort' (60,392) are more common than 'type' (10,669) in the spoken genre. All in all, the word 'type' is the most formal while other type words are more related to a colloquial language.

Table 2: The distribution of Type, Kind, Sort in different genres arranged by their frequencies

Type			Kind			Sort		
Genre	Frequency	Per Million	Genre	Frequency	Per Million	Genre	Frequency	Per Million
Academics	27911	233	Spoken	117536	931.82	Spoken	60392	478.79
Website	22222	178.84	TV	78627	613.92	Blog	20155	156.71
Blog	18944	147.29	Blog	47170	366.76	Fiction	19786	167.22
Magazine	15661	124.2	Fiction	45872	387.69	TV	18489	144.36
Spoken	10669	84.58	Website	41155	331.22	Website	16682	134.26
News	8945	73.48	Magazine	34751	275.6	Magazine	12433	98.6
TV	7613	59.44	News	31891	261.96	News	9710	79.76
Fiction	5415	45.76	Academics	19984	166.82	Academics	7202	60.12
Total	117380	118.2	Total	416986	419.92	Total	164849	166.01

Semantic Preferences

In response to the second research objective, the following subsections present the noun and adjective collocation in relation to their semantic preferences.

Noun collocates of type, kind and sort

The first type of collocations of these three words are nouns. Table 3 below shows the top-twenty noun collocates of these three words. The word 'type' can be collocated with diabetes, cell, cancer and so on while 'kind' can be used with stuff, trance, Shortland, limbo and others. Moreover, many nouns such as thing, compromise, nonsense, and epiphany could be used with the word 'sort'.

Table 3: Noun collocates of type, kind and sort

Rank	Type			Kind			Sort		
	Noun collocate	Frequency	MI value	Noun collocate	Frequency	MI value	Noun collocate	Frequency	MI value
1	diabetes	3276	6.97	stuff	3717	3.1	thing	11401	3.35
2	cell	1759	3.3	trance	69	3.48	compromise	173	3.53
3	cancer	1161	3.15	shorthand	67	3.67	nonsense	139	3.35
4	error	1145	4.06	limbo	55	3.01	epiphany	51	4.64
5	personality	984	4.26	craziness	52	3.58	homecoming	37	4.26
6	supernova	310	6.14	weirdo	46	3.22	prank	36	3.54
7	bacteria	238	3.24	hush	39	3.03	limbo	33	3.94
8	attribute	210	3.4	catharsis	29	3.79	mischief	31	3.72
9	fracture	197	4.02	silliness	27	3.01	trance	30	3.95
10	tumor	195	3.09	mix-up	26	3.53	pilgrimage	28	3.03
11	journalism	193	3.07	nihilism	25	3.23	goodies	28	3.4
12	vegetation	156	4.18	snobbery	24	3.8	consolation	26	3.09
13	biography	145	3.34	weirdness	23	3.08	truce	26	3.19
14	receptor	139	4.04	triage	23	3.1	litmus	24	4.48
15	diabetic	132	5.85	sicko	23	4.21	disclaimer	21	3.17
16	influenza	130	4.09	trickery	22	3.24	rebirth	21	3.21
17	font	124	4.09	demagoguery	22	3.81	commotion	18	3.04
18	insulin	112	3.7	elitism	19	3.05	fetish	18	3.49
19	severity	103	3.38	nastiness	19	3.3	shenanigans	18	3.75
20	neuron	98	3.13	fatalism	19	3.47	voodoo	17	3.03

As shown in Table 3, there are some noun collocations that can be used in both 'kind' and 'sort'. These words are 'trance' and 'limbo' as shown examples (1) to (4) below.

(1)Typically a Kvaedamadur would go into a kind of trance state for seven months of the year (just in the evenings if .....

(2)I was soon able to walk, turn, walk turn, with my eyes closed, so I would get into this like sort of trance...

(3)From the moment Robert Mueller was appointed special counsel, nearly two years ago, Washington has existed in a kind of political limbo. The White House ....

(4)The crew members themselves remain in a sort of legal limbo. The four Russians onboard were allowed to return to their homes ....

Semantic preferences of noun collocates of type

An analysis of semantic preferences based on their noun collocates of types found that there are eight types of categories (shown in Table 4). The first category is related to medicine. Examples of noun collocations of this category include diabetes, cell, cancer, bacteria. The second category refers to the Technology, and the words malware and cybersecurity. The third category highlights the term related to information exchange including journalism, annotation, inference. The fourth category is physiology, listing the collocations such as personality, maltreatment, infidelity, reinforcement while the fifth category is related to science e.g., integer, conductivity, subtype. More specifically, the noun collocations such as vegetation, maize, chondrite are categorized as botany. In the seventh category, social group consisted of words including celebrity and religiosity. Lastly, the miscellaneous category encompasses diverse terms from error to polarity. In order to see these noun collocates words in the context, examples (5) to (12) below show the concordance lines retrieved from COCA websites.

Table 4: Semantic preference of noun collocates of type

Category	Examples of Words
Medicine	diabetes, cell, cancer, bacteria, tumor, insulin, influenza, severity, neuron
Technology	malware, font, cybersecurity, firewall, virus, software, database, algorithm
Communication	journalism, annotation, inference, communication, language, disfluency,
Psychology	personality, infidelity, giftedness, maltreatment, relativity, reinforcement
Science	integer, conductivity, estimator, multilocus, luminosity, neutrino, subtype,
Botany	vegetation, maize, chondrite, sorghum, warts
Social group	celebrity, yuppie, religiosity, conscript, civilization, tradition
Miscellaneous	error, supernova, attribute, fracture, biography, finalist, specifications

(5).... grains have been associated with a lower risk of heart disease, diabetes and some types of cancer.

(6).... with the advice given in this article you should be able to remove any type of malware from your computer and get it back to working order.

(7).... and obscure compiler issues, maybe it's time to switch languages to something with type inference, closures, etc.

(8)....particular for the dwarf characters, is pretty much dependent on the personalities and the type of characters they are.

(9)just specified, the first capacitor electrode region semiconductor region is of the same conductivity type as the drain region.

(10).... for grassland bird conservation, it would be helpful to understand how variation in vegetation types and management affects grassland bird use of fields.

(11)Laurie felt that the public had to be warned, especially this group of yuppie types.

(12).... knowledge and are treated as though they are fully capable of setting specifications for the type of education they wish to receive.

Semantic preferences of noun collocates of kind

Table 5 presents the semantics preference analysis of noun collocates of the word 'kind' and it reveals three distinct themes. The first of them refers to emotions and states. These noun collocates signify several emotions such as trance, limbo, craziness, hush, mix-up and weirdness as illustrated in (13). The second theme, 'Communication,' includes nouns related to different forms of communication and interaction, such as shorthand, snobbery, trickery, demagoguery, elitism, and politicking, exemplified in (14). Lastly, the third theme, 'Relationships,' consists of nouns which refer to the different kinds of relationships such as stuff and coronet, with an instance of 'symbiosis' presented in (15).

Table 5: Semantic preference of noun collocates of kind

Categories	Examples of Words
Emotions and States	trance, limbo, craziness, hush, catharsis, silliness, mix-up, weirdness
Communication	shorthand, snobbery, trickery, demagoguery, elitism, politicking
Relationships	stuff, symbiosis, coronet

(13)The more I think that they're out there promoting this kind of craziness and silliness, the more that the American people are put of ….

(14)…. is completely irrelevant. This comment is, IMHO, an example of the worst kind of food snobbery.

(15)…. your essential passion becomes your job? When properly supervised the two should form a kind of symbiosis.

Semantic preference of noun collocates of sort

As displayed in Table 6 below, the semantic preference analysis of noun collocates of 'sort' presents eight themes. In the first theme, Emotion includes words of emotional states and experiences, including epiphany, catharsis, nostalgia, stupor, reverie, premonition, and weirdness, as evident in (16). The second theme of this semantic preference refers to social interactions, and it includes collocates such as compromise, homecoming, mischief, prank, truce, and commotion, as illustrated in (17). Additionally, the third theme is related to belief including voodoo, fetish, talisman, incantation, magic, occult, and enchantment, as demonstrated in (18). The fourth theme is associated with metaphysical concepts like dualism, nihilism, and equivalency in (19). The fifth theme or value encompasses nouns such as normalcy, censure, elitism, and demarcation, as seen in (20). The sixth theme refers to conflict such as compromise, truce, vendetta, altercation, catharsis, and contraption, as shown in (21). Science, the seventh theme, includes litmus and barometer in (22), while the last theme is legal category containing nouns such as malfeasance, defamation, and miscommunication in (23).

Table 6: Semantic preference of noun collocates of sort

Categories	Examples of Words
Emotions	epiphany, catharsis, nostalgia, stupor, reverie, premonition, weirdness
Social Interactions	compromise, homecoming, mischief, prank, truce, commotion
Belief	voodoo, fetish, talisman, incantation, magic, occult, enchantment
Metaphysical Concepts	dualism, nihilism, equivalency
Values	normalcy, censure, elitism, demarcation
Conflict	compromise, truce, vendetta, altercation, catharsis, contraption
Science	litmus, barometer
Legal	malfeasance, defamation, miscommunication

(16).... on the tile floor for the first but hardly last time -- I had a sort of epiphany.

(17).... It was a show I'd been looking forward to -- a homecoming of sorts.

(18)No, I've got to pick up my little sister. Was this some sort of sick fetish thing?

(19).... But do you see where there's the same thing, that there's a sort of moral equivalency here?

(20)Yet the difference between " turning the corner " and a return to any sort of normalcy is vast.

(21) There is no incentive because there's no reward to actually trying to broker any sort of compromise

(22).... weighty and important social movement behind them, but it shouldn't be a sort of litmus test.

(23)... away as fast as it had come. A cover-up? It sounded like some sort of boring malfeasance.

Adjective collocates of type, kind and sort

The second type of collocation of these three words is adjectives. As shown in Table 7, the top 20 results based on frequency and MI value (≥3) were presented with the adjective collocates. As seen in the Table, the most frequent collocates of type is different, followed by certain, various and others, respectively. Also, the most common collocate of kind is also different. However, the most frequency of sort is middling. Moreover, it should be noted that there are only 11 adjective collocates as the cut-off point criteria.

Table 7: Adjective collocates of type, kind and sort

Rank	Type			Kind			Sort		
	adjective collocate	Frequency	MI value	adjective collocate	Frequency	MI value	adjective collocate	Frequency	MI value
1	different	8503	3.87	different	9941	3.02	middling	85	7.09
2	certain	2779	3.69	just	2378	3.36	perverse	41	3.91
3	various	1878	3.75	imaginable	84	3.43	unspoken	21	3.07
4	specific	1344	3.24	considerate	69	3.94	platonic	16	3.64
5	particular	1116	3.1	marrying	58	7.15	negotiated	15	3.02
6	distinct	269	3.21	courteous	41	3.13	dreamy	15	3.07
7	movable	157	6.77	hearted	23	3.66	nefarious	14	3.38
8	spectral	111	5.03	fucked-up	21	3.44	hypnotic	11	3.06
9	hard-hitting	86	6.28	take-charge	20	5.34	good-natured	11	3.41
10	diabetic	62	4.48	in-your-face	19	3.39	godly	10	3.07
11	localized	54	4.24	glass-half-full	14	6.51	no-nonsense	10	3.09
12	predominant	43	3.77				zero-sum	10	3.79
13	adventurous	36	3.2				contemptible	9	4.18
14	moveable	33	5.91				catch-all	9	4.87
15	imaginable	30	3.02				messianic	8	3.12
16	blue-collar	28	3.13				macabre	8	3.24
17	marrying	26	7.07				self-fulfilling	8	3.32
18	alveolar	25	5.54				conspiratorial	7	3.02
19	neuronal	24	3.72				cathartic	7	3.14
20	outdoorsy	24	6.3				gregarious	7	3.14

Semantic preference of adjective collocates of type

Table 8 reports the four themes of the semantic preference analysis of adjective collocates of 'type'. The first category is quality which includes adjectives describing specific qualities and characteristics, such as different, certain, various, specific, particular, distinct, and predominant, as exemplified in (24). The second theme is medicine involving adjectives related to medical and biological contexts, such as Biology, diabetic, localized, alveolar, neuronal, epithelial, and endocrine, as seen in (25). For the third theme, descriptors include adjectives like movable, moveable, spectral, numeric, population-based, and user-defined, as presented in (26). Lastly, the category of personal traits introduces adjectives that describe personal char-

acteristics and traits, including blue-collar, marrying, sanitary, assistive, outdoorsy, nerdy, and arty, as illustrated in (27).

Table 8: Semantic preference of adjective collocates of type

Categories	Examples of Words
Quality	different, certain, various, specific, particular, distinct, predominant
Medicine	Biology, diabetic, localized, alveolar, neuronal, epithelial, endocrine,
Descriptors	movable, moveable, spectral, numeric, population-based, user-defined
Personal Traits	blue-collar, marrying, sanitary, assistive, outdoorsy, nerdy, arty

(24)Pork industry has followed the beef industry's lead and bestowed different names onto different types of pork chops

(25)NaP is included in both pacemaker and non-pacemaker neurons and therefore both neuronal types are affected by simulated TTX and RZ blockade.

(26).... or BFP, since one portion of dbx code may have to operate on both types of floating-point data.

(27)" Lou likes hard-working people and gritty guys who come to play. The blue-collar type of hockey player. "

Semantic preference of adjective collocates of kind

The semantic preference analysis of adjective collocates of 'kind' provides three categories (as shown in Table 9). The first category refers to comparison, and it encompasses adjectives used for comparison and to convey specific qualities, such as different, just, imaginable, considerate, courteous, and hearted, as illustrated in (28). Emotion is the third category which consisted of adjectives like fucked-up, in-your-face, and glass-half-full to express emotions and attitudes, as exemplified in (29). Apart from that, the last category is relationship. It is dealt with adjectives e.g., marrying and take-charge, as seen in (30).

Table 9: Semantic preference of adjective collocates of kind

Categories	Examples of Words
Comparison	different, just, imaginable, considerate, courteous, hearted
Emotion	fucked-up, in-your-face, glass-half-full
Relationship	marrying, take-charge

(28).... fricking North Pole. I'd been all over and had seen just about every kind of person imaginable.

(29)....you could try to do one now, but this is the last day. Kind of fucked-up timing.

(30)I really do... But I can't. I'm just not the marrying kind.

Semantic preference of adjective collocates of sort

According to the findings in Table 10, the analysis of adjective collocates of 'sort' introduces several distinct categories. Emotion is considered as the first category in which adjectives were used to describe emotional attributes, including dreamy, good-natured, cathartic, gregarious, dour, genial, and likeable, as exemplified in (31). The relationship category includes adjective collocates which are used to reflect various types of relationships such as platonic, negotiated, godly, catch-all, gentlemanly, quasi, and vicarious, as seen in (32). Furthermore, behavioral trait as the third category consisted of adjectives such as middling, good-natured, gregarious, dour, genial, likeable, and gentlemanly, which describe behavioral characteristics and traits, as depicted in (33). The fourth category is ambiguous quality such as perverse, unspoken, nefarious, ulterior, subliminal, quasi, and pseudo, as illustrated in (34). Furthermore, the adjective collocated with sort to mention the mental states are hypnotic, self-fulfilling, and conspiratorial. These adjectives are the fifth category and the example is shown in (35). The sixth category is ethical quality and the adjectives are godly, no-nonsense, zero-sum, contemptible, and legalistic, associated, as illustrated in (36). Additionally, the seventh category includes supernatural adjectives such as godly, messianic, and macabre pertain to supernatural and mythical contexts, as seen in (37). Finally, the artistic category involves adjectives including jazzy and plaintive, as demonstrated in (38).

Table 10: Semantic preference of adjective collocates of sort

Categories	Examples of Words
Emotion	dreamy, good-natured, cathartic, gregarious, dour, genial, likeable
Relationships	platonic, negotiated, godly, catch-all, gentlemanly, quasi, vicarious
Behavioral Trait	middling, good-natured, gregarious, dour, genial, likeable, gentlemanly
Ambiguous Quality	perverse, unspoken, nefarious, ulterior, subliminal, quasi, pseudo
Mental States	hypnotic, self-fulfilling, conspiratorial
Ethical Quality	godly, no-nonsense, zero-sum, contemptible, legalistic
Supernatural	godly, messianic, macabre
Artistic	jazzy, plaintive

(31) It kind of sort of happened on its own, but it was a sort of cathartic week for everybody involved.

(32).... is good and the players fun to hang with, but see old-school gaming as sort of the Platonic ideal of a gaming campaign.

(33).... ring that he had on his finger. And my mother is a very gregarious sort and just started engaging him in conversation.

(34).... great scientific revelations are there, but several vegans are active and I get some sort of perverse pleasure in taking aim at.

(35).... " Joseph placed an elbow on the bar and leaned over in a conspiratorial sort of way.

(36).... upon the cross of Christ will cause us to mourn for sin after a godly sort.

(37).... writers contribute regularly to daily publications. Together with the journalists, they feel a sort of a messianic vocation: They want to become a mouthpiece for the people.

(38)Tom Collins, the graduate teaching assistant and radical, is " a sort of jazzy, Tom Waitsy kind of guy.

Discussion

Once again, this study investigated the differences among three near synonyms e.g., type, kind and sort and examined their noun and adjective collocation in relation to semantic preferences. After the data collection and analysis process, it was found that this present study provides two major findings. The following paragraphs discuss each result.

The first result found that the frequency among these three words varies across genres in COCA. This could be suggested that in spite of their seminaries in overall meaning, the words type, kind and sort are not able to be used interchangeably. This result is consistent with other previous studies (Kruawong & Phoocharoensil, 2022; Lertcharoenwanich, 2023; Phoocharoensil, 2020b, 2021b; Phoocharoensil & Kanokpermpoon, 2021; Sridhanyarat & Phoocharoensil, 2023). As previous studies found different sets of near synonym could be classified and exemplified to the learners and users, this could facilitate them to use near synonym appropriately (Phoocharoensil, 2013).

Secondly, an analysis of collocates in relation to the semantic preference found that the word 'type' is mostly associated with medical contexts while 'kind' and 'sort' often appear in discussions about groupings and classifications. For the medical contexts, the explanation of use with medical terms could be explained by the fact that the word 'type' is also considered one the academic medical word list composed by Lei and Liu (2016). Additionally,

the words kind and sort seem to be more near synonym than the word 'type'. This study shares some similarities with previous studies (Lertcharoenwanich, 2023). It is possible that the meanings of two words could be closer to another word.

Suggestions

Based on the result of this study, this present study introduces the following suggestions:

1. For the pedagogical implication, the results suggest that Type, Kind and Sort are used differently despite their similar meaning. According to West (1953), these three words are common in the GSL list which is very frequent as the threshold to study the English language. Teaching this common vocabulary in a second language classroom could potentially help raise the awareness of learners and their comprehension (Nation, 2001; Schmitt & Schmitt, 2014). To this end, language teachers should teach the differences among these three words and other sets of words in the classroom because the learners could not differentiate the near synonyms (Phoocharoensil, 2021).

2. The second one is the teaching material. When it comes to implementing the research results in the classroom, the result of this study could be served as research-based evidence in the second language classroom. As Phoocharoensil (2020a, 2020b) suggested, this evidence should be used in the classroom to point out the differences among near synonyms and also illustrate how to use the reference corpus. As a consequence, language teachers could take advantage of the results in this study when they compose the teaching materials for L2 learners.

However, this study is limited based on the following aspects:

1. First, this study only focusses on three criteria of differentiating near synonyms. Therefore, more
studies are needed to explore the same set or the new set of near synonyms with other criteria to contribute new knowledge to this research area.

2. Second, the corpus of this research is very limited because it focuses on American English. According to Chaokongjakra (2023), Lertcharoenwanich (2023) and Lertcharoenwanich and Phoocharoensil (2022), there should be other reference corpora such as British National Corpus or BNC. It would be highly recommended that this exploration could possibly provide an insightful aspect of near synonym in British English.

3. Lastly, as mentioned in the data collection, the verb collocates under cut-off points of MI scores were excluded from this present study. For this reason, further research is

necessary to reduce the cut-off point of MI scores to allow results of verb collocates among these words.

References

Aimjirakul, P. (2015). **Big and Large as Near Synonyms.** Thammasat Review, 16(3), 1–15.

Ayhan Yavuz. (2014). **Materials Development for Adult Learners in Teaching Vocabulary.** Sino-US English Teaching, 11(4).

Biber, D. (2006). **University language: a corpus-based study of spoken and written registers**. In Studies in corpus linguistics.

Biber, D. (2009). **A corpus-driven approach to formulaic language in English**. International Journal of Corpus Linguistics, 14(3), 275–311.

Celen, K. M., & Yalçin, Ş. (2021). T**he Effects of Vocabulary Resources Use on Lexical Rich ness in L2 Writing**. Milli Egitim, 50(230).

Chaokongjakra, W. (2023). **Is it 'important,' 'significant,' or 'crucial'? A Corpus Based Study of English Synonyms**. LEARN Journal: Language Education and Acquisition Research Network, 16(2).

Chung, S.-F. (2011). **A Corpus-based Analysis of "Create" and "Produce**." Chang Gung Jour nal of Humanities and Social Sciences, 4(2).

Daller, H., Milton, J., & Treffers-Daller, J. (2007). **Modelling and Assessing Vocabulary Knowl edge**. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), Modelling and Assessing Vocabulary Knowledge. Cambridge University Press. https://doi.org/10.1017/ CBO9780511667268

Fang, L., Ma, Q., & Yan, J. (2021). **The effectiveness of corpus-based training on collocation use in L2 writing for Chinese senior secondary school students**. Journal of China Computer-Assisted Language Learning, 1(1).

Gablasova, D., Brezina, V., & McEnery, T. (2017). **Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence**. In Language Learning (Vol. 67).

Jirananthiporn, S. (2018). **Is this problem giving you trouble? A corpus-based examination of the differences between the nouns problem and trouble**. Thoughts, 2, 1–25.

Kamali Khalavi, S., & Zeraatpishe, M. (2023). **Relationship of Iranian EFL Learners' Vocabulary Depth with Their Writing Vocabulary Use, Fluency, and Organization**. Journal of Contemporary Language Research, 2(1).

KOÇAK, A. (2020). **EFL Students' Perceptions of Using COCA to Develop Their Vocabulary**. Journal of Language Research, 4(1).

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).

Laufer, B. (1990). **Ease and Difficulty in Vocabulary Learning: Some Teaching Implications**. Foreign Language Annals, 23(2).

Laufer, B. (1997). **The Lexical Plight in Second Language Reading**. In J. Coady & T. Huckin (Eds.), Second Language Vocabulary Acquisition (pp. 20-34). Cambridge University Press.

Lei, L., & Liu, D. (2016). **A new medical academic word list: A corpus-based study with en hanced methodology**. Journal of English for Academic Purposes, 22.

Lertcharoenwanich, P. (2023). **Analysis of Collocations and Semantic Preference of the Near-synonyms: Blank, Empty, and Vacant.** LEARN Journal: Language Education and Acquisition Research Network, 16(1).

Lertcharoenwanich, P., & Phoocharoensil, S. (2022). **A Corpus-based Study of the Near-syn onyms: Purpose, Goal and Objective**. REFLections, 29(1).

Milton, J. (2009). **Measuring second language vocabulary acquisition**. In Measuring Second Language Vocabulary Acquisition. https://doi.org/10.1080/09571736.2012.677623

Milton, J., & Hopwood, O. (2022). **Vocabulary in the Foreign Language Curriculum**. In Vocabu lary in the Foreign Language Curriculum. https://doi.org/10.4324/9781003278771

Nation, I. S. P. (2001). **Learning Vocabulary in Another Language**. Cambridge University Press. Petcharat, N., & Phoocharoensil, S. (2020). A Corpus-Based Study of English Syn onyms: Chance and Opportunity. REFLections, 27(2), 218–245.

Phoocharoensil, S. (2013). **Cross-linguistic influence: Its impact on L2 English collocation production**. English Language Teaching, 6(1).

Phoocharoensil, S. (2020a). **A Genre and Collocational Analysis of Consequence, Result, and Outcome**. 3L: Language, Linguistics, Literature, 26(3).

Phoocharoensil, S. (2020b). **Collocational patterns of the near-synonyms error, fault, and mistake**. International Journal of Communication and Linguistic Studies, 19(1).

Phoocharoensil, S. (2021a). **Multiword units and synonymy: Interface between collocations, colligations, and semantic prosody**. GEMA Online Journal of Language Studies, 21(2).

Phoocharoensil, S. (2021b). **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., & Kanokpermpoon, M. (2021). **Distinguishing the near-synonyms 'in**

crease' and 'rise': Genre and collocation investigation**. Kasetsart Journal of Social Sciences, 42(4).

Schmitt, N., & Schmitt, D. (2014). **A Frequency Dictionary of English: Core Vocabulary for Learners**. Routledge.

Sridhanyarat, K., & Phoocharoensil, S. (2023). **A CORPUS-BASED INVESTIGATION OF ENGLISH NEAR-SYNONYMS: ASSESS, EVALUATE, AND MEASURE**. Humanities, Arts and Social Sciences Studies, 23(1).

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

West, M. (1953). **A general service list of English words**. London: Longman, Green.

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

Yaacob, A., Shapii, A., Saad Alobaisy, A., Al-Rahmi, W. M., Al-Dheleai, Y. M., Yahaya, N., & Alamri, M. M. (2019). **Vocabulary Learning Strategies Through Secondary Students at Saudi School in Malaysia**. SAGE Open, 9(1)