



Exploring Near-Synonymy of the Adjectives ‘Expensive’ and ‘Costly’ through Corpus-Based Study

Kanjanapa Intarajak¹, Theerapong Kongduang¹, Clement Robles Bana¹

¹Kasetsart University, Thailand.

ARTICLE INFO

Article history:

Received May 15, 2025

Revised June 16, 2025

Accepted June 30, 2025

Keywords:

Near-Synonymy,
Adjectives,
Corpus-Based Study,
Formality,
Collocational Patterns,
Semantic Preferences

Conflict of Interest:

None

Funding:

None

ABSTRACT

This study employed a corpus-based methodology to investigate the near-synonymous English adjectives ‘expensive’ and ‘costly’, with particular attention to their degree of formality, collocational patterns, and semantic preferences. Drawing on data from the Corpus of Contemporary American English (COCA), which encompasses eight distinct genres (spoken, fiction, magazines, newspapers, academic texts, blogs, web pages, and TV/movie subtitles), the analysis focused on the frequency of use across genres and the most common noun collocates with Mutual Information (MI) scores of 3 or higher. The results indicated that ‘expensive’ occurred more frequently in informal genres such as blogs, magazines, and spoken texts, while ‘costly’ was predominantly found in formal contexts, especially newspapers and academic writing. Although some collocates overlapped, their semantic preferences differed: ‘expensive’ typically co-occurred with terms related to luxury products and personal choices, while ‘costly’ was associated with business, legal, and problem-related contexts. The findings revealed that the two adjectives, while semantically related, were not fully interchangeable. This distinction provided insight into how English learners could develop vocabulary precision through the use of corpus tools. The study concluded that corpus-based analysis was an effective method for highlighting subtle lexical differences, offering practical implications for language teaching and learning.

Corresponding Author: Theerapong Kongduang, Department of Hospitality Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University, No.1, Kamphaeng Saen, Nakhon Pathom, 73140, Thailand. Tel. +66-94-2514691. E-mail: Theerapong.k@ku.th.



© Kanjanapa Intarajak, Theerapong Kongduang, Clement Robles Bana

This is an open access article under the CC BY-SA 4.0 international license

1. Introduction

1.1 Background of Study

Vocabulary is accepted as essential to communicating or conveying messages successfully; as Wilkins (1972) explained, “Without grammar, very little can be conveyed; without vocabulary, nothing can be conveyed.” However, mastering English vocabulary can be challenging and problematic for English language learners because there are many synonyms, words that share similar or close meanings with other words in the same language.

Phoocharoensil and Kanokpermpoon (2021) assert that identifying synonymous words in English is a major obstacle for foreign language learners, primarily because of the challenge of distinguishing their semantic

nuances. According to Laufer (1990), learners consistently face difficulties in vocabulary learning due to the presence of synonymy. Due to the unique meanings of each synonym and their dependence on language context, it is impossible to employ them interchangeably in all situations.

Within the numerous words in synonym sets, two words—*expensive* and *costly*—were focused on in this researcher's interest. Both words are in the top 3000 and 5000 important words English learners need to know aligned to the Common European Framework of Reference for Languages (CEFR) (Oxford Learner's Dictionary, 2024), implying that these words are common and that learners need to understand how they are used in context.

Furthermore, As stated in the Longman Dictionary of Contemporary English (LDOCE), *expensive* is a typical English term with high frequency, ranking among the top 1,000 most spoken and 2000 most written words. While *costly* is one of the terms ranked at medium frequency, it is practical and valuable for English learners to look at how native English speakers use them in context because both adjectives are often employed in spoken and written English. In particular, language learners benefit from exposure to authentic materials (Timmis, 2015). As a result, it is reasonable to expect that studying informal words will help learners approach a language more successfully and use words more accurately.

1.2 Research Questions

- 1) What is the degree of formality of the synonyms *expensive* and *costly* based on the distribution across different genres?
- 2) What are the similarities and differences between the synonyms *expensive* and *costly* regarding collocation and semantic preference?

1.3 Definition of Terms

- 1) **COCA** stands for the Corpus of Contemporary American English. It is a massive collection of spoken and written texts representing how English is used in real-life situations. These texts are categorized into eight genres: magazines, blogs, newspapers, webpages, spoken language, fiction, academic texts, and TV and movie subtitles.
- 2) **Corpora/Corpus (plural)** is an extensive and organized database of texts stored in electronic form. It is considered a tool for linguists and researchers to study and analyze language patterns and usage across different disciplines.
- 3) **Synonyms** are words in the same language that share the same or a very close meaning. By way of illustration, 'big' and 'large' are considered synonyms because they convey the idea of something being considerable in size.

2. Literature Review

This section will discuss three aspects of the relevant academic literature: the theoretical concepts of corpus linguistics and synonyms and previous research on English synonyms through a corpus-based study.

2.1 Corpus Linguistics

According to Sinclair (1991), word meaning cannot be conveyed by a single word itself but rather by a combination of words in sequence. This idea is the cornerstone for corpus linguistics, which means computer-based collections of authentic text.

Historically, a corpus is defined as a single author's work, such as the Shakespeare corpus. In 1961, W. Nelson Francis and Henry Kucera of Brown University began creating the Brown Corpus, the first computerized collection of English texts for linguistic research.

The definition of a corpus by O'Keeffe et al. (2007) as large and structured collections of spoken and written texts in digital form stored on a computer gathered from various sources and accessible for both qualitative and quantitative assessment by linguists and researchers to study language patterns and usage across disciplines.

2.1.1 Characteristics of the Corpus Approach

Biber, Conrad, and Reppen (1998) identified four characteristics of the corpus approach. Firstly, it is practical because it examines real-life language usage trends in authentic settings, reflecting everyday situations where linguistic communication occurs. Secondly, the data are drawn from a huge collection of naturally existing texts for a foundation of analysis. The corpus may be classified as spoken, written, or academic. Thirdly, the corpus approach involves the computer as an essential tool in collecting and analyzing language in a corpus. This indicates that the corpus approach requires the involvement of computers. Fourth, the corpus approach is concerned with two main areas of analysis: quantitative analysis and qualitative

analysis. The quantitative data obtained from the corpus is subsequently employed in a qualitative way to respond to research inquiries.

2.1.2 Types of Corpora

Eight categories of corpora were identified by Lindquist (2009): spoken corpora, specialized corpora, general corpora, historical corpora, multilingual and parallel corpora, text archive corpora, and web corpora.

Bennett (2010) gave a different way to classify corpora that can be useful in the classroom. This classification includes general corpora, specialized corpora, learner corpora, educational corpora, historical corpora, parallel corpora, comparable corpora, and monitor corpora. Some examples of general corpora that collect written and spoken language from various sources are the American National Corpus (ANC), the British National Corpus (BNC), and the Corpus of Contemporary American English (COCA). These gather language from news articles, journals, books, scholarly writings, laws, and business contexts.

Additionally, O'Keeffe et al. (2010) divided corpora into three main types: monolingual corpora, comparable corpora, and parallel corpora.

2.2 Synonym Theory

The term *synonym* originates from the Greek roots. *syn* means 'alike' and *onym* means 'name', used to describe words with similar or nearly identical meanings. (Murphy, 2010).

Oxford Advanced Learner's Dictionary, 9th edition (OALD, 2015) identifies a synonym as "a word or expression that has the same or nearly the same meaning as another in the same language" (p. 1589). Moreover, the Longman Dictionary of Contemporary English, 6th edition (2014) provides a similar meaning: "a word with the same meaning as another word in the same language" (p. 1862), e.g., "shut is a synonym of closed."

2.2.1 Types of Synonyms

Synonyms are typically divided into two classifications, referred to by different names: "strict, perfect, or absolute synonyms" (Taylor, 2002) and "loose or near-synonyms" (Cruse, 1986).

Strict synonyms are two words employed equally in all contexts without changing meaning, style, or connotation. Chung (2011) said it is uncommon to find strict synonyms in language usage since words cannot always be used interchangeably in all contexts. A point supported by Phoocharoensil (2020) explained that this synonymy is rare because of its potential to create redundancy in language.

Strict synonyms can be found in the English language, such as '*everyone*' and '*everybody*.' Jarunwaraphan and Mallikamas (2020) demonstrated how the words '*everyone*' and '*everybody*' can be used interchangeably in every context. Everyone (every person) has to drink to survive, and everybody (every person) has to drink to survive. The '*everyone*' and '*everybody*.' from the previous example does not affect the meaning of the words, even if there is a word change in the sentence.

Loose synonyms refer to words with similar meanings but cannot be used alternately in every context. The words '*find*' and '*discover*' are often confusing. "Marie Curie discovered radium in 1898," the word '*found*' cannot be substituted for 'discovered' because '*discover*' refers to being the first encounter with something, whereas '*find*' indicates "experience something in some way."

2.2.2 Criteria for Distinguishing Synonyms

Generally, numerous criteria are used to distinguish synonyms. However, the level of formality, collocations, and semantic preference as the main focus areas in this study in approaching the distinction between *expensive* and *costly* adjectives.

Degree of Formality

Phoocharoensil and Petcharat (2017) highlighted that synonyms may not be suitable in all contexts, whereas synonyms have a comparable core meaning. In particular, many words are commonly used in formal settings, for example, academic articles or conferences, while others are used in informal settings, e.g., everyday conversations. Examples of synonyms differentiated by their degree of formality include the following:

Formal	Informal
ask	inquire/enquire
finish	conclude
help	assist
teach	instruct

(Longman Online Dictionary of Contemporary English)

Moreover, Leech and Svartvik (2002) provided another criterion for defining the character of formal and informal language.

Formal language has the following characteristics:

- Including words originating from French, Latin, and Greek.
- Being more precise than a casual conversation.
- Refrain from using pronouns such as I, you, and we.
- Avoid addressing the writers or the readers directly.
- Having its own grammatical characteristics.

Informal language has the following characteristics:

- Including words originating from Anglo-Saxon.
- Using imprecise phrases, idioms, or colloquial words.
- Using personal language
- Comprising contracted verb forms.
- Comprising phrasal verbs.

Collocations

The definition of collocation can be found in the Cambridge Online Dictionary (2024). “a word or phrase that is often used with another word or phrase in a way that sounds correct to people who have spoken the language all their lives but might not be expected from the meaning.”

Words within a set of synonyms often have the ability to appear along with several alternative collocates. According to Palmer (1997), many synonyms have differences in the restrictions of their collocations. For example, *acquire* and *obtain* share some noun collocates, but they may not be replaced interchangeably in every context. *Obtain* appears to collocate with a noun relating to money, which does not occur with the word *acquire*. (Sittironnarit et al., 2022).

Semantic Preference

Semantic preference refers to the tendency of words to form extended units of meaning by associating closely with other semantically related words (Begagić 2013). It is also known as semantic association, which refers to a word frequently associating closely with a specific collocation (Flowerdew, 2012). McEnery and Hardie (2012) also recommend grouping words based on their lexical identification when identifying semantic preference. For instance, the word *cause* is often followed by other terms referring to diseases, e.g., cancer and disease.

2.3 Previous Corpus-Based Studies on English Synonyms

Over the years, numerous researchers have analyzed the data acquired from various corpus software, namely, the Corpus of Contemporary American English (COCA) or the British National Corpus (BNC), to identify the similarities and differences in a set of synonyms. Below are several prior study studies on English synonyms.

Chung (2011) analysed similarities and differences in *create* and *produce* using the meanings of these two synonyms and the semantic components of their respective vocabularies. The analysis was based on three corpora (Brown Corpus, Frown Corpus from ICAME) and the BNC (British National Corpus). The two synonymous verbs had overlapping meanings. It is interesting to note that the semantic implications of the types of objects seem to influence the choice of the two verbs (e.g. quantity). *Produce* is associated with things that represent large quantities.

Thamratana (2013) researched similarities and differences between five synonyms: *reduce*, *decrease*, *diminish*, *dwindle*, and *decline*, based on their meanings, degree of formality, and grammatical patterns. The data was extracted from two dictionaries and the Corpus of Contemporary American English (COCA). The word *decrease* tends to be more frequent in academic texts, while *dwindle* tends to be more frequent in magazines. Additionally, the finding showed that five synonyms have the same fundamental meaning but differ in word collocates and grammatical structures. It was determined that the five synonyms, *reduce*, *decrease*, *diminish*, *dwindle*, and *decline*, cannot be used interchangeably in every situation.

Jarunwarapha and Mallikamas (2020) used the Corpus of Contemporary American English (COCA) and online dictionary data to analyze how *chance* and *opportunity* are used in academic texts. According to the

study, opportunity is frequently employed in academic contexts, whereas chance is less commonly used in academic and spoken texts. It has been discovered that near-synonyms may show distinctive semantic preferences. As for the grammatical pattern, target synonyms coincide with prepositions of, to, and for. In addition, the semantic preferences of *chance* and its associates are lower than those of *opportunity*, and nearly synonymous terms show distinctive collocation and semantic prosody despite having common semantic similarities.

Petcharat and Phoocharoensil (2017) explored the meanings, degrees of formality, collocations, and grammatical patterns of *appropriate*, *proper*, and *suitable* adjective synonyms. The data was obtained by combining information from three dictionaries with the Corpus of Contemporary American English (COCA). The result suggests that *appropriate* is more formal than *suitable* and *proper*. Moreover, they share a basic meaning, but their specific meanings differ. For instance, *suitable* can describe a person, while *appropriate* cannot. These three synonyms share a limited number of collocations, and grammatical patterns lead to uninterchangeable use in all instances.

Bunnag (2021) three synonyms were studied for their meaning, formalities, and collocations. These synonyms are *expensive*, *costly*, and *pricey*. Bunnag consulted three online dictionaries for their definitions and combined them with the Corpus of Contemporary American English (COCA) to analyze the probability of the occurrence of these synonymous terms across eight distinct genres. The study revealed that the three synonyms share a similar meaning of requiring a large amount of money and that they are frequently used in magazines, newspapers, and blogs. However, there is only one word, 'alternative', in the collocations of these adjectives.

Ajmal et al. (2022) compared the similarities and differences between the three synonyms *beautiful*, *handsome*, and *pretty* by their meanings and collocations through the Longman Dictionary of Contemporary English 6th edition (2014) and the British National Corpus (BNC). The result indicates that three synonyms, *beautiful*, *handsome*, and *pretty*, have a common core meaning but differ in their detailed meanings and collocation. For example, *beautiful* is appropriate to use with women, while *handsome* is for men. Moreover, among the three synonyms, *handsome* has more collocates than *beautiful* and *pretty*, but The word *beautiful* has the most frequent noun collocates compared to the other two synonyms.

Sittironnarit et al. (2022) analyzed the genre and word combinations of verbs, *acquire* and *obtain*, utilizing the Corpus of Contemporary American English (COCA). Both verbs exhibit a significant level of formality and are more commonly used in written forms of communication rather than in spoken language. The noun collocates that were shared by only five. English learners ought to thoroughly contemplate the noun collocate when utilizing these words. For instance, the source is a word that commonly occurs with the verb *obtain*, but it does not typically occur with the verb *acquire*. While they have some similar semantic preferences, these two terms have certain nouns that commonly occur with them.

Although the main focus of this study may align with Bunnag (2021), the author would like to find additional insight into these two adjectives: *expensive* and *costly*. Exploring the degree of formality across eight different genres and examining how these synonyms are similar and how they are different regarding collocation and semantic preference remains valuable for understanding language variation and usage.

3. Method

The following part provides a comprehensive overview of the research methodology employed in the study, including details on the research sample, procedures, data collection, and data analysis.

3.1 Sample

The primary objective of this study was to analyze the distribution of various genres, collocations, and semantic preferences to differentiate between *expensive* and *costly*. The corpus data of these synonyms was searched on the Corpus of Contemporary American English (COCA).

COCA, the corpus of American English provided, is the most extensive and accessible collection covering various genres and representing different writing and speaking styles. It was used to analyze synonyms for *expensive* and *costly*. It was divided into eight genres which ensures the possibility of generalizing and relying on data: spoken, fiction, magazines, newspapers, academic texts, TV and movie subtitles, blogs, and web pages (Davies, 2020). Consequently, it examines how the target words and their collocates are distributed across various genres. In addition, COCA is a supervised corpus that consistently updates its data to ensure validity. It encompasses a remarkable one billion data, comprising approximately 25 million words gathered between 1990 and 2020.

Additionally, COCA is particularly different from other corpora, focusing on the top 60,000 words and providing information for each word. The display in COCA has multiple features that offer frequency data,

definitions, synonyms, WordNet entries, related topics, concordances, clusters, keywords, and KWIC/concordance lines. These make COCA ideal for researchers, teachers, and learners.

3.2 Data Collection Procedure

To address the first research question, the Corpus of Contemporary American English (COCA) was utilized to analyze the degree of formality of the synonym *expensive* and *costly*, based on the distribution across eight genres, which are spoken, fiction, magazines, newspapers, academic texts, TV and movie subtitles, blogs, and web pages. The researcher explored <https://www.english-corpora.org/coca/>. Then, the researcher pressed the 'chart' function on the menu bar and entered *expensive* and *costly* synonyms to find their frequency independently.

To address the second research question about noun collocations, the researcher accessed free software from <https://www.english-corpora.org/coca/>. The researcher selected the 'word' function from the menu bar and inputted synonyms for the terms *expensive* and *costly*. Subsequently, the author selected the 'collocates' function from the menu. The synonyms were categorized based on their respective parts of speech. In particular, the target words of this study are adjectives; the most common noun collocates were concentrated according to the collocational strength rated by the Mutual Information (MI) score or value.

The MI score was utilized, as stated by Gablasova, Brezina, and McEnery (2017) to determine how often the target word and the collocation occurred together. On the other hand, collocational association results cannot be accurately predicted using only the MI score because it typically signifies a high score for uncommon combinations (Cheng, 2012). It can be implied that collocations with high MI scores may appear with low frequency, which means they cannot be used in the analysis because they do not represent strong collocations. For this study, The MI score is frequently utilized to measure the co-occurrence strength between a target adjective and noun, indicating their collocation Schmitt (2010). Therefore, the examination of the top-twenty noun collocations with the highest frequency and the MI score level (≥ 3) (Cheng, 2012) of the target adjective was collected from COCA.

The following stage identified twenty noun collocates of each adjective, *expensive* and *costly*, depending on their semantic preference. As stated in another way, semantically similar collocations were categorized according to the same theme. This process of semantic classification enhances comprehension of the distinctions among the synonyms regarding how they are selected alongside other words and in what contexts they appear.

3.3 Data Analysis

The framework inspired by Phoocharoensil (2020) was employed to examine the data of this study. The COCA data was categorized into two groups and analyzed to answer the above research questions. Data distribution was analyzed across eight genres, utilizing total frequency and frequency within each genre. The nouns listed in each target word's top twenty most frequent were identified based on frequency and MI score value (≥ 3). Then, Similar-sounding noun collocates were assigned to the same category based on semantic preference, allowing for an obvious distinction between target near-synonyms.

4. Results and Discussion

This part responds to the two research questions. It discusses significant points drawn from the findings and analysis, covering the frequency and distribution on the whole and in each genre of the synonyms *expensive* and *costly*. Also, the similarities and differences between noun collocates and semantic preferences are covered.

Table 1. Meanings and Usage of the Synonyms *Expensive* and *Costly* from LDOCE

	<i>expensive</i>	<i>costly</i>
meaning	1. costing a lot of money	1. very expensive, especially wasting a lot of money 2. something that is costly causes a lot of problems or trouble
sentence examples	1. She spends most of her money on <i>expensive</i> clothes. 2. My uncle took us out to dinner at an <i>expensive</i> restaurant. (Longman, 2024)	1. Buying all those computers was a <i>costly</i> mistake. 2. A lawsuit would be <i>costly</i> and we would probably lose. (Longman, 2024)

Table 1 provides the definitions of the synonyms *expensive* and *costly* searched from the Longman Dictionary of Contemporary English Online (2024) to distinguish between expensive and costly. It shows that the definitions of *expensive* and *costly*, requiring a large sum of money, appear comparable. Nevertheless, whereas *costly* can relate to losses or sacrifices that money alone may not be able to make whole, *expensive* seems more closely linked to material values.

Furthermore, it appears that the term *costly* describes the practice of paying for something that is not worthwhile and impending issues or problems that might or might not be resolved with money alone.

The following table provides the result of the first study question; their frequency distribution across different genres determines the degree of formality of the *costly* and *expensive* synonyms.

Table 2. Distribution of the Synonyms *Expensive* and *Costly* across Genres according to Frequency of Occurrence

Rank	<i>expensive</i>			<i>costly</i>		
	Genre	Frequency	Per million	Genre	Frequency	Per million
1.	magazines	9,713	77.03	newspapers	2,442	20.06
2.	blogs	8,924	69.39	magazines	2,425	19.23
3.	newspapers	8,366	68.72	academic texts	2,132	17.80
4.	webpages	7,174	57.74	blogs	1,503	11.69
5.	spoken	4,846	38.42	webpages	1,442	11.61
6.	fiction	4,473	37.08	spoken	672	5.33
7.	academic texts	3,966	33.11	fiction	321	2.71
8.	TV and movies subtitles	3,328	25.98	TV and movies subtitles	143	1.12
	Total	50,790	51.15	Total	11,080	11.16

Table 2 shows that *expensive* (50,790 tokens) appears to be used more frequently than *costly* (11,080 tokens) in every genre. However, when considering specific genres, the frequency per million of *expensive* is highest in magazines, blogs, newspapers, webpages, spoken language, fiction, academic texts, and TV and movie subtitles, respectively. It can be inferred that *expensive* commonly occur in informal rather than formal contexts. Conversely, *costly* exists in the genres of newspapers, magazines, and academic texts as the top three highest ranks.

The findings are directly in line with previous findings by Bunnag (2021). Notably, the genres with the highest occurrence of the word *costly* are newspapers, magazines, and academic texts, all of which are written materials. Although *expensive* is more commonly used in academic texts than *costly*, the frequency of *expensive* in the academic texts genre is the second lowest among all the genres. Therefore, it can be confirmed that *costly* occurs more frequently in formal contexts than *expensive*.

The following tables address the similarities and differences between noun collocates and semantic preferences of *expensive* and *costly* in response to the second research question.

Table 3. Noun Collocates of *Expensive* and *Costly* from COCA

Rank	<i>Expensive</i>			<i>costly</i>		
	Noun collocates	Frequency	MI Score	Noun collocates	Frequency	MI Score
1.	<u>equipment</u>	453	4.44	mistake	221	4.97
2.	suit	311	3.78	regulation	113	4.45
3.	restaurant	263	3.08	treatment	91	3.05
4.	clothes	260	3.44	<u>litigation</u>	88	6.64
5.	wine	209	3.15	battle	81	3.35
6.	<u>proposition</u>	191	5.01	<u>alternative</u>	76	4.47
7.	<u>alternative</u>	166	3.48	<u>equipment</u>	63	3.72

8.	taste	148	3.20	investment	62	3.04
9.	jewelry	145	4.78	disaster	58	3.99
10.	toy	140	3.48	procedure	52	3.22
11.	watch	137	3.25	error	50	3.44
12.	clothing	106	3.38	<u>repair</u>	49	5.05
13.	hardware	95	3.61	penalty	48	4.13
14.	electricity	90	3.08	lawsuit	46	4.08
15.	<u>repair</u>	73	3.50	delay	45	4.53
16.	leather	71	3.09	hurricane	42	3.84
17.	gasoline	71	3.75	turnover	38	5.49
18.	<u>litigation</u>	69	4.18	<u>proposition</u>	31	4.50
19.	perfume	68	4.71	endeavor	30	5.13
20.	hobby	59	3.95	infrastructure	26	3.45

The twenty nouns collocate with MI scores ≥ 3 were obtained from COCA and ordered based on frequency. Table 3 illustrates the noun collocations of these adjectives, *expensive* and *costly*. This study supports evidence from previous research (Bunnag, 2021); notably, the researcher found five nouns - *equipment*, *proposition*, *repair*, *litigation*, and *alternative* - share collocates with both adjectives. Therefore, based on the shared collocates of both adjectives, it can be concluded that they have both similar and different noun collocates.

After extracting the noun collocates of the words *expensive* and *costly* from COCA using the aforementioned approaches, all the nouns were categorized into distinct groups as follows:

Table 4. Semantic Preference of Noun Collocates of *Expensive*

Semantic Categories	Noun Collocates
1. PRODUCT	<i>clothes, clothing, leather, perfume, suit, toy, watch, wine, jewelry</i>
2. PERSONAL PREFERENCE	<i>alternative, hobby, taste</i>
3. ENERGY	<i>electricity, gasoline</i>
4. EQUIPMENT	<i>equipment, hardware</i>
5. PLACE	<i>restaurant</i>
6. BUSINESS	<i>proposition</i>
7. REMEDY	<i>repair</i>
8. LAW	<i>litigation</i>

Although the following results are slightly different to some extent considerably from those of Bunnag (2021), some discussions have shared the same argument.

Table 4 illustrates how the noun collocates of the word *expensive* can be categorized into eight semantic categories: PRODUCT, PERSONAL PREFERENCES, ENERGY, EQUIPMENT, PLACE, BUSINESS, REMEDY and LAW. In Bunnag's study, nouns were divided into only five groups. Both studies found a similar finding: the PRODUCT category contains the highest number of noun collocates. The term *expensive* is generally used to describe the high monetary value associated with items commonly regarded as luxurious, such as leather, perfume, suits, watches, and jewelry. In addition, the most striking result that emerges from the eight different semantic categories is that *expensive* mostly collocates with nouns, which are often used in informal contexts such as PRODUCT and PERSONAL PREFERENCES.

The discussion can be applied to confirm and reinforce the response to research question one, which claims that the word *expensive* is more frequently used in informal situations than *costly*.

Table 5. Semantic Preference of Noun Collocates of *Costly*

Semantic Categories	Noun Collocates
1. BUSINESS	<i>endeavor, infrastructure, investment, procedure, proposition, turnover</i>

2. LAW	<i>battle, lawsuit, litigation, penalty, regulation</i>
3. PROBLEM	<i>delay, disaster, error, hurrican, mistake</i>
4. REMEDY	<i>repair, treatment</i>
5. EQUIPMENT	<i>equipment</i>
6. PERSONAL PREFERENCE	<i>alternative</i>

Table 5 demonstrates that the noun collocates of the word can be categorized into six distinct semantic groups: BUSINESS, LAW, PROBLEM, REMEDY, EQUIPMENT, and PERSONAL PREFERENCE. The biggest semantic category in this study is BUSINESS, totaling six words. This outcome contradicts Bunnag (2021), who found the highest number of noun collocates *costly* in the group of LEGAL PROCESSES. The semantic categories of *costly* frequently appear with noun collocates, which are often used in formal contexts such as BUSINESS and LAW, it might be utilized as additional proof to support the conclusion of the first research question. Moreover, the meanings of *costly* from The Longman Online Dictionary of Contemporary English (2024) confirm that it refers to something that causes many issues, falling under the semantic category of PROBLEM.

In addition, it is interesting to note that five nouns concatenate with both adjectives in Table 3, except for the word alternative. The word equipment, proposition, repair, and litigation fall under the semantic category of BUSINESS and LAW. It suggests that both *expensive* and *costly* can be used in formal contexts such as BUSINESS or LAW. However, considering the observed shared collocates, *costly* may be more suitable because these words are highly associated.

5. Conclusion

The main objective of this study is to examine the frequency and distribution on the whole and in each genre of the synonyms, *expensive* and *costly*, and the similarities and differences between noun collocates and semantic preferences. Sources for the corpus data include the Corpus of Contemporary American English (COCA) as the main instrument. Considering the result of the study, it can be concluded that *expensive* and *costly* cannot be used interchangeably. The study indicates that the definitions of *expensive* and *costly*, requiring a large sum of money, appear comparable. Nevertheless, whereas *costly* can relate to losses or sacrifices that money alone may not be able to make whole, *expensive* seems more closely linked to material values.

Regarding formality, *expensive* is more commonly used in academic texts than *costly*, but the frequency of *expensive* in the academic texts genre is the second lowest among all the genres. Therefore, it can be confirmed that *costly* occurs more frequently in formal contexts than *expensive*. In terms of noun collocation, there are five nouns - equipment, proposition, repair, litigation, and alternative – found to be a collocation of both adjectives.

Based on the semantic category, the adjective *expensive* collocates with nouns in eight categories: product, personal preferences, energy, equipment, place, business, remedy, and law. The product category has the highest number of collocates, primarily used for luxury items. The semantic categories of *costly* often include noun collocates, often used in formal contexts like business and law. The meanings of *costly* from The Longman Online Dictionary of Contemporary English (2024) confirm its meaning as a problem causing numerous issues.

Moreover, this study has demonstrated that corpora can serve as an excellent resource for teaching English language. Students should be instructed to know the difference between words in a group of synonyms and should be taught to look up material in learners' dictionaries and corpus databases. Consequently, students can differentiate synonyms within a collection by utilizing corpora and learner's dictionaries as their primary educational resources.

It should be noted that this study focuses on the formalities, collocations, and semantic preference of two synonyms: *expensive* and *costly*. It suggests including dialects and grammatical patterns in future research to understand synonyms better. The study focused on American English corpora, which may not apply to other English varieties like British English. Future research should include data from dictionaries and corpora from other dialects, such as British, Canadian, or New Zealand. Moreover, the top-20 lists were used to limit the number of noun collocates, but expanding the investigation to include lower frequency collocates will provide more accurate collocational patterns for the target synonyms.

Acknowledgements

The authors wish to express sincere gratitude to the Dean of the Faculty of Hospitality Industry, Kasetsart University, for generous institutional support and for fostering a research-friendly environment. Special thanks are also extended to the Programme Chair of English for Intercultural and Entrepreneurial

Communication for his continuous encouragement, valuable guidance, and academic mentorship throughout the research process. Appreciation is also given to those who assisted with proofreading and offered constructive feedback during the development of this study.

References

- Ajmal, M., Kumar, T., Ritonga, M., & Nukapangu, V. (2022). A Corpus-Based Analysis of the Adjectives and Synonyms - Beautiful, Handsome, and Pretty., 12(2), 159.
- Begagić, M. (2013). Semantic preference and semantic prosody of the collocation make sense. *Jezikoslovlje*, 14, 403-416.
- Bennett, G. R. (2010). *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Michigan ELT. Retrieved from <https://www.press.umich.edu/pdf/9780472033850-part1.pdf>
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. New York: Cambridge University Press.
- Bunnag, N. (2021). *A corpus-based study of English synonyms: expensive, costly, and pricey*. Thammasat University.
- Cheng, W. (2012). *Exploring corpus linguistics: Language in action*. Routledge.
- Chung, S. (2011). A corpus-based analysis of “create” and “produce”. *Chang Gung Journal of Humanities and Social Sciences*, 4(2), 399-425.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
- Flowerdew, L. (2012). *Corpora and language education*. Springer.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*, 67, 155-179.
- Jackson, H., & Amvela, E. (2000). *Words, meaning and vocabulary: Meaning relations*. Great Britain: Atheneum Press.
- Jarunwaraphan, B. , & Mallikamas, P. (2020) . A corpus- based study of English synonyms: chance and opportunity. *rEFlections*, 27(2), 218-245. <https://files.eric.ed.gov/fulltext/EJ1283247.pdf>
- Laufer, B. (1990). Words you know: How they affect the words you learn. In J. Fisiak (Ed.), *Further insights into contrastive linguistics* (pp. 573–593). Amsterdam: JohnBenjamins.
- Leech, G. N., & Svartvik, J. (2002). *A communicative grammar of English*. London: Longman
- Lindquist H., (2009). *Corpus linguistics and the description of English*. Croydon,England: Edinburgh University Press.
- Longman Dictionary of Contemporary English*. (2014). Essex: Pearson Education.
- Longman Dictionary of Contemporary English. (2024). *Longman Dictionary of Contemporary English*, Retrieved from <http://www.ldoceonline.com/>
- McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Murphy, M. L. (2010). *Lexical meaning*. Cambridge, United Kingdom: Cambridge University Press.
- O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge University Press.
- O'Keeffe, A. & McCarthy, M.(2010) Historical perspective: What are corpora and how have they evolved?, in O'Keeffe, A. & McCarthy, M.(eds), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, p 3-13.
- Oxford Advanced Learner's Dictionary*. (2015). Oxford: Oxford University Press.
- Oxford Learner's Dictionary. (2024). *Oxford Advanced Learner's Dictionary*, Retrieved from <http://www.oxforddictionaries.com/>
- Palmer, F. R. (1997). *Semantics*. Great Britain: Cambridge University Press.
- Petcharat, N., & Phoocharoensil, S. (2017). A corpus-based study of English synonyms: appropriate, proper, and suitable. *LEARN Journal: Language Education and Acquisition Research Network Journal*, 10(2), 10-24.
- Phoocharoensil, S. (2010). A corpus-based study of English synonyms. *International Journal of Arts and Sciences*, 3(10), 227-245.

Phoocharoensil, S., & Kanokpermpoon, M. (2021). Distinguishing the near-synonyms 'increase' and 'rise': Genre and collocation investigation. *Kasetsart Journal of Social Sciences*, 42(4), 968 – 975. <https://so04.tci-thaijo.org/index.php/kjss/article/view/255757/173077>

Sinclair, J. (1991). *Corpus, concordance and collocation*. Oxford: Oxford University Press.

Sittironnarit, S., Khunasathitchai, K., Kosashunhanan, K., & Kumdee, S. (2022). Acorpus-based analysis of English synonyms: acquire and obtain. *Journal of Roi Kaensarn Academi*, 7(9), 242-252.

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

Thamratana, P. (2556). *A corpus-based analysis of reduce decrease diminish dwindle and decline*. Thammasat University.

Timmis, I. (2015). *Corpus Linguistics for ELT: Research and Practice (1st ed.)*. Routledge. <https://doi.org/10.4324/9781315715537>

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

The Authors:

Kanjanapa Intarajak (<https://orcid.org/0009-0007-4886-6152>), Programme Committee of English for Intercultural and Entrepreneurial Communication, Department of Hospitality Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University. Her research interests include English for Specific Purposes, English Language Teaching, Corpus Linguistics, Linguistics, Intercultural Communication. Email: Kanjanapa.in@ku.th.

Theerapong Kongduang (<https://orcid.org/0009-0009-0961-6147>), Programme Chair of English for Intercultural and Entrepreneurial Communication, Department of Hospitality Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University & Vice Dean for Academic Affairs and Public Relations, Faculty of Hospitality Industry, Kasetsart University. His research interests include English Language Teaching, Testing & Assessment, Intercultural Communication, Syntax, Pragmatics, Discourse Analysis, Linguistics. E-mail: Theerapong.k@ku.th

Clement Robles Bana (<https://orcid.org/0009-0006-3973-8127>), Programme Committee of English for Intercultural and Entrepreneurial Communication, Department of Hospitality Innovation and Intercultural Communication, Faculty of Hospitality Industry, Kasetsart University. His research interests include English Language Teaching, Intercultural Communication, Tourism and Hospitality Management. E-mail: Clementrobles.b@ku.th