Exploring Collocational Patterns and Genres: An Analysis of the Use of *Poisonous* and *Venomous* in American English

Supakorn Phoocharoensil[1]

[1] Language Institute, Thammasat University, Bangkok, Thailand

Correspondence: Supakorn Phoocharoensil, Language Institute, Thammasat University, Bangkok, Thailand. Tel: +66897999323. E-mail: supakorn.p@litu.tu.ac.th

Received: April 25, 2024 Accepted: September 5, 2024 Online Published: December 12, 2024

doi:10.5430/wjel.v15n2p202 URL: https://doi.org/10.5430/wjel.v15n2p202

Abstract

This research study explores the usage differences between the synonymous adjectives *poisonous* and *venomous*. Based on the Corpus of Contemporary American English (COCA), comprising over one billion words, representing American English from 1990 to 2023 across eight genres, i.e., spoken language, fiction, magazines, newspapers, academic texts, webpages, weblogs, and TV/movie subtitles, this study investigates the distribution of both synonyms across genres and their collocational patterns. This study addresses: 1) the distribution of *poisonous* and *venomous* across genres, and 2) common noun collocates of these synonyms. The findings reveal both adjectives are prevalent in informal contexts, such as fiction and magazines. However, distinct collocational preferences emerge in that *poisonous* aligns with nouns related to gases, chemicals, substances, and food, while *venomous* is usually combined with nouns denoting actions, emotions, and food items. This study highlights the interplay among lexical choice, formality, and collocational preferences in English.

Keywords: synonym, corpora, genre, collocation, collocational patterns

1. Introduction

1.1 Introduction

Vocabulary, often compared to "the fuel of language, without which nothing meaningful can be understood or communicated" (Gardner, 2013, p. 2), holds a pivotal role in English language teaching, with a discernible correlation between vocabulary proficiency and the development of learners' core language skills (Nation, 2013). A central obstacle encountered by English learners in acquiring vocabulary in their second language is synonymy, denoting the linguistic phenomenon where multiple words share identical meanings (Carter, 2012). As demonstrated by Phoocharoensil (2020a), despite semantic resemblances between words (e.g., *consequence* vs. *outcome*), substituting one for the other can impact the natural flow of English production in a second language. This effect arises due to specific word combinations (e.g., *strong wind*) forming widely recognized collocational patterns, unlike others (e.g., **strong rain*). Students with limited exposure to their second language may struggle to discern the most suitable synonyms for specific contexts.

Numerous studies have concentrated on distinguishing English near-synonyms across various linguistic dimensions, including formality levels (e.g., Boontam & Phoocharoensil, 2022; Chaengchenkit, 2023; Chaokongjakra, 2023; Jirananthiporn, 2018; Narkprom, 2024; Phoocharoensil, 2020b, 2021a, 2021b; Sridhanyarat & Phoocharoensil, 2023), connotations (e.g., Phoocharoensil, 2020a; Stubbs, 1995), collocations (e.g., Boontam & Phoocharoensil, 2022; Chaengchenkit, 2023; Chaokongjakra, 2023; Crawford & Csomay, 2016; Jirananthiporn, 218; Narkprom, 2024; Phoocharoensil, 2020a, 2020b, 2021a, 2021b; Sridhanyarat & Phoocharoensil, 2023), semantic prosody (e.g., Phoocharoensil, 2021a, 2021b; Selmistraitis, 2020), and colligations (e.g., Phoocharoensil, 2021a, 2021b).

This study aims to investigate the usage of the adjectives *poisonous* and *venomous*, which are often confusing due to their similarity, by analyzing their presence across different types of texts and their common collocations. By using extensive and authoritative American English data sources such as COCA, the research seeks to provide a clearer understanding of how each adjective is appropriately applied. The meanings and common collocations of these words were carefully examined using the Longman Dictionary of Contemporary English (2014) to confirm that they are true synonyms. Following this, the research explored how these words appear in various text genres within COCA and identified the nouns that frequently accompany them, offering a deeper comparison of the themes associated with these collocations.

1.2 Literature Review

This section provides an overview of corpus linguistics, synonymy, and the methods for discerning synonyms using different criteria.

1.2.1 Corpus Linguistics and Vocabulary Teaching

Corpora are vast collections of texts found in natural language contexts (Biber, Conrad & Reppen, 1998). Language corpora are large and structured collections of written or spoken texts used for linguistic research and analysis. They are used as a resource for investigating

language use, identifying patterns, and testing hypotheses about language (Friginal, 2018). Cheng describes corpus linguistics as "the compilation and analysis of corpora" (2012, p. 6). Concordance lines are essential for using corpus data to study language patterns. Lindquist and Levin define a concordance as a list of word contexts in a given text (2018, p. 5). Corpus-based and corpus-driven analyses typically show data as keyword-in-context (KWIC) concordances, where the target term is referred to as a node.

Corpus linguistics offers benefits to researchers, ELT practitioners, and language learners by ensuring authenticity, reliability, and speed (Lindquist & Levin, 2018). Corpus linguistics provides a novel approach to analyzing language use that goes beyond native speaker intuition (Hunston, 2002). According to Conrad (2010), corpus data can reveal common and uncommon language choices, as well as typical and untypical patterns within a context. Corpus linguistics examines authentic data patterns using frequency-based analysis (Szudarski 2018). Frequency-based evidence can be considered more trustworthy than intuition alone. It is important to note that corpus analysis can be undertaken in both qualitative and quantitative ways (O'Keeffe, McCarthy & Carter, 2007). However, while numerical statistics, such as frequency and MI scores, are automatically provided through a corpus-based approach, human judgement is still necessary to evaluate the given data by examining concordances or co-text to improve understanding of the data (Conrad, 2010).

1.2.2 Synonymy

Synonymy is a critical concept within lexicology and language pedagogy. Carter (2012) defines synonymy as a "symmetrical sense relation where multiple linguistic forms can be said to share the same conceptual or propositional meaning" (p. 34). For example, the terms *house, home, abode*, and *domicile* all refer to 'the place where someone resides' (Carter, 2012). Synonyms are generally categorized into two main types: perfect synonyms and near-synonyms. As Cruse (1986) describes, 'perfect synonyms' or 'absolute synonyms' are words with entirely identical meanings, permitting their interchange in any context without affecting the original meaning, style, or connotation. However, such absolute synonyms are exceedingly rare or non-existent, as Jackson and Amvela (2007) note, leading to linguistic redundancy.

Conversely, 'loose synonyms', also known as 'near-synonyms', are words with similar but not identical meanings, where the degree of semantic overlap can vary depending on the context (Phoocharoensil, 2020a, 2020b). Jackson and Amvela (2007) emphasize that near-synonyms are not always contextually interchangeable. For example, it is permissible to use the near-synonyms *repair* and *mend* interchangeably in (1); however, in (2), where the context involves clothing, *mend* sounds more appropriate than *repair.* In English Language Teaching (ELT), the focus is typically on varying degrees of near-synonyms rather than absolute synonyms.

(1) I will *mend/repair* that light in the hall.

(2) My father used to *mend/*repair* his shoes.

(Phoocharoensil, 2020a, p. 3)

The ability to differentiate between near-synonyms is crucial for both ELT instructors and learners. Despite their similar denotative meanings, the usage of these words can vary, potentially causing confusion among English speakers, including non-native English-speaking teachers (Phoocharoensil, 2020a). Context plays a pivotal role in selecting the appropriate synonym for a given situation (Carter, 2012; Murphy, 2009). Linguists use specific criteria to distinguish near-synonyms, some of which will be examined in the following subsection.

1.2.3 Criteria for Distinguishing Synonyms

Lexicologists, who study the generalizations and regularities of word forms and their interrelations, employ several criteria to illustrate how near-synonyms are used differently across contexts (Jackson & Amvela, 2007). One criterion is connotation. Although two words may have similar core meanings, they can carry distinct associative or emotional nuances (Jackson & Amvela, 2007). For instance, although *clever* and *cunning* both pertain to an individual's ingenuity, *clever* carries a distinctly positive connotation, meaning "quick at learning and understanding," as illustrated in (3). In contrast, *cunning* conveys a negative connotation, implying the ability to achieve something through deceit, as shown in (4).

(3) Lucy is quite *clever* and does well at school.

(Longman Dictionary of Contemporary English 2014)

(4) He is as *cunning* as a fox.

(Oxford Advanced Learner’s dictionary 2014)

Another distinguishing criterion is the level of formality or style. Synonyms may differ in formality, with one being more formal and the other more casual. For instance, while *let* and *allow* are synonymous, the latter is generally used in more formal contexts compared to caveat (LDOCE, 2014). Corpus research has explored formality levels in using synonyms. Jirananthiporn (2018), for example, found that the noun *problem* is associated with more formal texts in COCA, whereas its synonym *trouble* appears more frequently in informal genres. Similarly, Jarunwaraphan and Mallikamas (2020) identified genre-based differences in the usage of the synonyms *chance* and *opportunity*. Phoocharoensil (2020a) systematically examined the similarities and distinctions among the synonyms *consequence, result,* and *outcome,* focusing on eight text types in the COCA corpus. The findings demonstrate that these synonyms are most frequently employed in academic texts, suggesting a strong association with more formal registers. More specifically, *consequence* and *outcome* exhibit their lowest frequency of use in TV and movie subtitles, while *result* occurs least frequently in fiction, both genres being characteristic of

informal English. Phoochaorensil (2020b), focusing on the usage distinctions among the synonyms *error*, *fault*, and *mistake*, yielded notable insights. *Mistake* appears with the highest frequency in COCA, followed by *fault* and *error*, respectively. In terms of formality, *error* is most prevalent in academic texts, affirming its association with a higher level of formality, whereas *mistake* and *fault* occur most frequently in informal contexts, such as TV and movie subtitles. Phoocharoensil (2021a) found that *persist* and *persevere* are commonly used in written English, as they occur more frequently in magazines, webpages, blogs, and academic texts—genres predominantly characterized by written language. Corpus data further suggests that *persist* and *persevere* are less prevalent in colloquial English, as evidenced by their relatively lower frequency in spoken language, fiction, and TV and movie subtitles, which represent more informal contexts. Their infrequent use in spoken genres supports their association with a high degree of formality, though *persevere* is slightly less formal, as indicated by its third-place ranking in academic texts compared to *persist*.

Collocations also play a significant role in distinguishing near-synonyms. Baker, Hardie, and McEnery (2006) define "collocation" as "the phenomenon where certain words are more likely to occur in combination with other words in specific contexts" (p. 36). Collocations are typically analyzed using statistical methods, such as mutual information (MI) scores (Saito, 2020). Murphy (2009) notes, "words tend to pattern with limited ranges of other words" (p. 156), suggesting that while near-synonyms may share conceptual meanings, they differ in their typical collocational patterns. For example, although *close* and *near* are similar in meaning, they collocate uniquely with different nouns, as in *close friends* and *near future*, as opposed to *close future* and *near friends*. These collocational constraints limit the interchangeability of synonyms across all contexts. Thornbury (2002) emphasizes that even minor modifications to collocations through the substitution of near-synonyms can result in the text deviating from standard usage (p. 7).

Examining typical collocations for a word requires consulting extensive native-speaker corpora. Murphy's (1998) study demonstrated that collocations are instrumental in assessing the degree of similarity between near-synonyms. For example, while *big* and *enormous* are contextually interchangeable, the two synonyms exhibit different collocational patterns despite overlapping meanings. This supports Szudarski's (2018) assertion that "no two words can be considered perfect synonyms, as corpus data reveal significant differences in phraseological patterns" (p. 43).

Recent research has utilized corpus methods to analyze collocations and elucidate differences in near-synonym usage. Several studies have focused on synonymous nouns (e.g., Jarunwaraphan & Mallikamas, 2020; Phoocharoensil, 2020a, 2020b, 2021a). For example, Jarunwaraphan and Mallikamas (2020) investigated the frequent word partners of *opportunity* and *chance* in COCA, finding that *opportunity* typically pairs with verbs conveying positive connotations, while *chance* often associates with verbs suggesting negative outcomes. Similarly, *opportunity* tends to collocate with adjectives linked to positive contexts, whereas *chance* pairs with adjectives that have mixed connotations due to its polysemous nature.

Recent studies using the updated version of COCA, which now includes eight genres (Davies, 2020), have furthered the analysis of typical collocations. Phoocharoensil (2020a) investigated the prevalent verb and adjective collocates of *consequence*, *result*, and *outcome*. A thorough examination of the semantic tendencies related to verb and adjective collocations reveals subtle variations in usage. The analysis shows that *consequence* is primarily associated with terms that carry negative connotations. In contrast, although *result* is a near-synonym of *consequence*, it does not strongly connect with negative or adversarial contexts; rather, its collocates are more commonly found in the realm of research methodology. Unlike *consequence* and *result*, *outcome* is characterized by a wider range of semantic collocates, with a recurring focus on topics or categories, exemplified by adjectives such as *psychological*. Phoocharoensil (2020b), focusing on the synonyms *error*, *fault*, and *mistake*, differentiated these synonyms through an analysis of their collocates. Specifically, *error* has the most extensive range of verb collocates, while *fault* is associated with the fewest verb collocates. Adjectives frequently collocated with *fault* are also the least numerous. Notably, the semantic preferences of these collocations reveal significant patterns. *Error* and *mistake* share some overlapping collocates within the same thematic groups, particularly those indicating seriousness—such as *fatal*, *glaring*, *grave*, and *grievous*—as well as those related to correcting or addressing the issue, like *correct* and *rectify*.

Phoocharoensil (2021a) conducted a study on synonymous verb differentiation, examining *persist* and *persevere*. Despite sharing the cognitive meaning of 'to continue doing something in a difficult situation,' the verbs are associated with distinct contexts, as evidenced by their differing sets of noun collocates. *Persist* tends to collocate with nouns denoting negative meanings, while *persevere* often combines with Christian-oriented vocabulary and phraseological units that reflect determination. Another study by Kruawong and Phoocharoensil (2022) focused on synonymous verbs *teach*, *educate*, and instruct, emphasizing object noun collocates. *Teach* is frequently associated with school or university subjects, while *educate* involves societal interaction and social work. *Instruct* often collocates with legal English terms. Chaengchenkit (2023), also analyzing collocational patterns of the synonymous verbs *cease*, *halt*, and *stop*, found that regarding the verb *cease*, most of its noun collocates fall into three major categories: ‘war’, ‘conflict’, and ‘damage’. Conversely, for *halt* and *stop*, the majority of their noun collocates can be classified into the categories ‘movement’ and ‘change’, and ‘emotion’ and ‘action’, respectively. In a similar way, Narkprom (2024) investigated the common collocates of two synonymous verbs, i.e. *restrict* and *constrain*. It was revealed that *restrict* demonstrates a more limited semantic scope of noun collocates compared to *constrain*. Both near-synonyms exhibit a pattern in which they co-occur with two sets of right-sided noun collocates, each containing entirely opposite morphological meanings. For example, *acc-* in *access* as seen in *restrict access*, contrasts with *ex-* in *expansion* as seen in *constrain expansion*. This observed pattern could greatly influence how ELT practitioners teach these two synonymous verbs.

In addition to examining synonymous nouns and verbs, certain studies have also investigated adjective synonyms. When analyzing the

common noun collocates of *whole*, *entire*, and *total*, Imsa-ard and Phoocharoensil (2022) observed that *whole* and *entire* share more common noun collocates, suggesting a higher degree of synonymity compared to *total*. The only noun commonly associated with both *entire* and *total* is *population*, while *whole* and *total* do not share any top-30 noun collocates. Similarly, Chaokongjakra (2023) utilized collocations to distinguish synonymous adjectives, revealing that *significant* tends to be linked with concepts related to quantity and downward trends, whereas *crucial* is more commonly associated with the political realm. Examinations of collocations between nouns and adverbs showed both shared and distinct preferences among synonymous adjectives. *Important* often collocates with adverbs indicating intensity, *significant* with those related to academic fields, and *crucial* with adverbs that express unique cultural themes. Additionally, unique noun and adverb collocates were identified, highlighting the individual semantic associations of each synonymous adjective.

Although there have been corpus-based studies distinguishing usage differences between synonymous adjectives (e.g., Chaokongjakra, 2023; Murphy, 1998; Phoocharoensil, 2022), it seems that no study so far has investigated the near-synonyms *poisonous* and *venomous* using the updated function in COCA known as *Word*, which shows the distribution of the search words across genres and also identifies common collocates with which they are used. The present study employs two major criteria—degree of formality across genres and collocational patterns—to differentiate between *poisonous* and *venomous*. The reasons why both words were selected lie in the fact that *poisonous* is classified as a B1 word based on the CEFR, which is worth teaching to English learners, while *venomous* is the near-synonym of *poisonous* identified by the Oxford Advanced Learner's Dictionary. This study aims to address the following research questions:

1. What are the core meanings of the synonyms *poisonous* and *venomous*?

2. What is the distribution of the synonyms *poisonous* and *venomous* across various genres?

3. Which nouns commonly co-occur with the synonyms *poisonous* and *venomous*?

2. Method

2.1 Data Collection

For this study, the Corpus of Contemporary American English (COCA) served as the primary resource. COCA stands as one of the most widely used corpora for representing American English. It encompasses over one billion words of text, with approximately 25 million words added annually from 1990 to 2023. This corpus covers eight genres, including spoken language, fiction, popular magazines, newspapers, academic texts, TV/movie subtitles, blogs, and other web pages, as of March 2020 (Davies, 2020). To be more specific, the information on the data source of each genre is provided below:

TV/Movie subtitles: Contains 128 million words from American TV and movie subtitles sourced from OpenSubtitles. The language is informal, similar to or more colloquial than actual spoken conversations.

Spoken: Comprises 127 million words transcribed from unscripted conversations on over 150 TV and radio programs such as NPR's All Things Considered, PBS's Newshour, and shows like Good Morning America and 60 Minutes. Also includes talk shows like Hannity and Colmes and Jerry Springer.

Fiction: Consists of 120 million words from short stories, plays, literary magazines, children's magazines, popular magazines, and movie scripts published from 1990 to the present.

Popular Magazines: Includes 127 million words from nearly 100 different magazines covering various domains like news, health, home and gardening, women's interests, finance, religion, and sports. It provides a wide variety across different years.

Newspapers: Comprises 123 million words from ten US newspapers such as USA Today, New York Times, Atlanta Journal Constitution, and San Francisco Chronicle. Covers sections, such as local news, opinion, sports, finance, etc.

Academic Journals: Contains 121 million words from nearly 100 peer-reviewed journals across the Library of Congress classification system. Covers a wide range of academic subjects.

Blogs: Includes 125 million words from blogs in the GloWbE corpus, representing a subset of US-based blogs. Originally sourced when Google allowed searches restricted to blogs.

Web pages: Contain 130 million words and represents a subset of "General" texts from the United States in the GloWbE corpus, including blogs and other web content.

Alongside other significant corpora forming part of Brigham Young University's 'mega-corpora,' COCA has been extensively employed in both research and English Language Teaching (ELT) for several reasons (Friginal, 2018). Firstly, researchers can explore word frequencies across various well-balanced genres to analyze common collocates and compare lexical usage among genres. Secondly, as a continually updated corpus, COCA provides a reliable source of contemporary native-speaker English, aiding ELT practitioners in staying informed of the present use of English. Moreover, COCA can facilitate students' autonomous and inductive learning through a corpus-based method, such as data-driven learning (DDL), which presents authentic English data to learners who are to make observations on the usage or rules based on the set of data to which they are exposed (Yamtui & Phoocharoensil, 2019).

Given these advantages, COCA was chosen for distinguishing the near-synonyms *poisonous* and *venomous*. Not only is COCA an

enormous corpus of English but also COCA, similar to other corpus software programs like AntConc or WordSmith Tools, acts as a corpus tool allowing learners to conduct a word search to examine the language patterns in which they are interested, e.g.,, collocations. With the introduction of *Word*, a new feature in the latest version of COCA, users gain access to essential information about searched words, such as their distribution across genres, definitions, related topics, collocates, etc. This feature is available for words ranked among the top 60,000 in COCA (Davies, 2020). As noted by Ma and Mei (2021), COCA enables users to understand nuanced meanings and usage patterns of keywords and related words, facilitating the systematic differentiation of near-synonyms based on typical collocates and associated genres.

2.2 Data Analysis

The current study aims to address the two aforementioned research questions. Initially, COCA was consulted to determine the frequencies and genre distributions of the target synonyms (i.e. *poisonous* and *venomous*) across the eight genres. Subsequently, the study explored noun collocates, i.e., nouns that often appear in combination with the target synonymous adjectives, frequently accompanying these synonyms. Collocates were extracted based on their collocational strength, considering both Mutual Information (MI) scores and frequency. High MI scores, while indicative of strong collocational associations, were not the sole criterion, as they might prioritize rare combinations not encountered in everyday language. Therefore, collocates were selected based on frequency, ensuring common and recurrent associations were prioritized. The top 25 high-frequency noun collocates in COCA with MI scores ≥ 3, signifying significant collocational strength, were listed.

Next, the common noun collocates of *poisonous* and *venomous* were categorized according to semantic preference, grouping collocates with similar meanings. The study then examined collocates frequently co-occurring with one or more synonyms, indicating potential semantic overlap. Additionally, nouns specifically co-selecting with certain synonyms were highlighted to illustrate how each synonym associates with particular collocates.

3. Results and Discussion

Table 1. The definitions of *poisonous* and *venomous* from three learner dictionaries

Synonym	Oxford Advanced Learner's Dictionary	Longman Dictionary of Contemporary English	Cambridge Dictionary
poisonous	1) causing death or illness if swallowed or taken into the body e.g., *poisonous* chemicals/fumes/plants e.g., a *poisonous* substance e.g., This gas is highly *poisonous*. 2) (of animals and insects) producing a poison that can cause death or illness if the animal or insect bites you SYNONYM *venomous* e.g., *poisonous* snakes e.g., a tiny spider with a *poisonous* bite	1) containing poison or producing poison e.g., Some mushrooms are extremely *poisonous*. e.g., *poisonous* gases such as hydrogen sulfide e.g., poisonous substances e.g., She was bitten on the ankle by a *poisonous* snake. 2) full of bad and unfriendly feelings e.g., the poisonous atmosphere of the office 3) *British English* someone who is poisonous seems to get pleasure from causing arguments, unhappiness etc e.g., That *poisonous* bastard Lucett told Morris I was seeing his wife.	1) very harmful and able to cause illness or death: e.g., *poisonous* chemicals e.g., Can you tell the difference between *poisonous* mushrooms and edible varieties? 2) A poisonous animal or insect uses poison in order to defend itself: e.g., a *poisonous* snake
venomous	1) (of a snake, etc.) producing venom e.g., a highly venomous snake	1) full of hatred or anger e.g., Lisa shot him a *venomous* glance. e.g., Reid reserved his most	1) A venomous snake, insect, etc., produces venom (= a poisonous liquid that can be put into another animal's body by biting or stinging):

	2) (formal) full of bitter feelings or hate e.g., a *venomous* look e.g., a *venomous* attack on his political enemies	venomous attack for the Rail Authority. 2) a *venomous* snake, insect etc produces poison SYN *poisonous*	e.g., Possible hazards include mosquitoes, sandflies, and *venomous* snakes. e.g., The Florida cottonmouth snake, which is also known as the water moccasin, has *venomous* fangs. e.g., She found a highly *venomous* black widow spider in a bunch of grapes.

A comparison of the two adjectives reveals their close similarity in meaning, as they both encompass two primary senses: 'causing illness or death' and 'the production of poison by animals or insects,' as demonstrated in entries from the Oxford Advanced Learner's Dictionary and the Cambridge Dictionary. Furthermore, the Longman Dictionary of Contemporary English expands on this by including two additional senses: 'being replete with negative and unfriendly emotions' and 'finding pleasure in provoking arguments or unhappiness,' both of which are closely linked.

Regarding *venomous*, all three dictionaries define it as a characteristic of animals, particularly snakes, capable of producing poison. Notably, both the *Oxford Advanced Learner's Dictionary* and the *Longman Dictionary of Contemporary English* classify *poisonous* and *venomous* as synonyms. Interestingly, these dictionaries also offer an alternative definition for both terms, denoting 'full of hate or bitter feelings'. From the dictionary entries, it is evident that these two adjectives are akin in meaning, posing a challenge for learners due to their shared major senses.

Table 2. Distribution of the synonyms *poisonous* and *venomous* across genres according to frequency

poisonous			***venomous***		
Genre	**Frequency**	**Per million**	**Genre**	**Frequency**	**Per million**
fiction	659	5.57	magazines	195	1.55
magazines	628	4.98	fiction	156	1.32
TV and movies subtitles	461	3.60	newspapers	112	0.92
webpage	405	3.26	webpage	100	0.80
weblogs	390	3.03	weblogs	100	0.78
newspapers	281	2.31	TV and movies subtitles	73	0.57
spoken	275	2.18	academic texts	49	0.41
academic texts	228	1.90	spoken	36	0.29
Total	3,327		Total	821	

According to COCA data, occurrences of *poisonous* (3,327 tokens) are approximately four times more frequent than those of *venomous* (821 tokens). Table 2 indicates that both synonyms are commonly found in informal written English. Specifically, *poisonous* appears most frequently in fiction (5.57 per million) and magazines (4.98 per million), respectively, while *venomous* is most prevalent in magazines (1.55 per million) and fiction (1.32 per million). This suggests a lower degree of formality associated with these synonymous adjectives, as evidenced by their infrequent appearance in academic texts, the most formal genre available in COCA. In academic texts, *poisonous* occurs only 1.90 times per million and *venomous* 0.41 times per million. The genres in which *poisonous* and *venomous* are least common are academic writings and spoken language, respectively.

Table 3. Noun collocates of poisonous and venomous in COCA

Rank	Noun collocate	*Poisonous* Frequency	MI Score	Noun collocate	*Venomous* Frequency	MI Score
1	snake	289	9.13	snake	168	10.43
2	gas	168	6.06	spider	28	8.75
3	plant	163	5.68	bite	26	7.86
4	substance	56	5.93	creature	15	6.05
5	atmosphere	54	5.88	insect	15	7.18
6	spider	53	7.61	snakebite	14	13.30
7	animal	45	3.75	animal	13	4.03
8	mushroom	42	7.10	species	12	4.69

9	chemical	41	5.99	spine	12	7.45
10	fruit	36	4.86	fang	12	9.17
11	tree	32	3.14	attack	11	3.74
12	berry	28	7.31	serpent	11	9.13
13	species	25	3.69	sting	10	8.52
14	bite	24	5.68	sea	9	4.02
15	duck	23	5.43	reptile	9	8.84
16	cloud	22	4.12	cobra	9	9.06
17	fume	21	8.02	viper	9	9.89
18	venom	21	8.15	hatred	8	6.60
19	food	20	8.35	toad	8	9.10
20	vapor	19	7.15	debate	6	3.30
21	seed	18	4.10	pit	6	5.78
22	human	17	3.23	reply	6	6.00
23	insect	17	5.29	ant	6	6.68
24	mercury	17	6.96	lizard	6	7.71
25	flower	16	3.61	scorpion	6	8.75

The nouns significantly collocating with the adjectives *poisonous* and *venomous*, determined by a mutual information (MI) value of ≥ 3, are listed in Table 3. Given their close core meanings as near-synonyms, these adjectives share some common noun collocates, such as *snake, spider, animal, species, insect*, and *bite*, which suggests their synonymous nature (Phoocharoensil, 2020a, 2020b). However, it is essential to carefully interpret the corpus-based data in Table 3 when considering these common collocations. For instance, this study's limitation is evident as only the top 25 nouns were targeted for collocation extraction. Other nouns may also co-occur with the target adjectives but are not listed in Table 3 due to their low frequency or MI score in COCA. For example, despite their high MI scores, common collocations like *poisonous snakebite* (MI score 9.43) and *venomous lionfish* (MI score 11.29) were not selected because their frequency did not rank among the top 25. Conversely, although *poisonous air* ranks seventh in COCA with 42 tokens, its MI score of 2.90 led to its exclusion from the collocation list.

Table 4. Semantic preference of noun collocates of *poisonous*

Categories	Collocates
1. CREATURE	animal, duck, human, insect, snake, species, spider
2. PLANT	berry, flower, fruit, mushroom, plant, tree, seed
3. GAS	atmosphere, cloud, fume, gas, vapor
4. CHEMICAL	chemical, mercury
5. SUBSTANCE	substance, venom
6. ATTACK	bite
7. FOOD	food

Following a comprehensive examination of the contextual associations of the adjective *poisonous*, seven primary categories emerged, as outlined in Table 4. The initial category, labeled CREATURE, encompasses seven collocates, namely *animal, duck, human, insect, snake, species*, and *spider*, as illustrated in (5). The second category, PLANT, consists of noun collocates referring to various plants and plant components, including *berry, flower, fruit, mushroom, plant, tree*, and *seed*, as depicted in (6). The third category, GAS, comprises collocates related to atmospheric substances such as *atmosphere, cloud, fume, gas*, and *vapor*, as exemplified in (7). The fourth and fifth categories, CHEMICAL and SUBSTANCE, each comprise two members: *chemical* and *mercury* in CHEMICAL, as shown in (8), and *substance* and *venom* in SUBSTANCE, as depicted in (9). The final two categories contain only one collocate each: "bite" in the "ATTACK" category and "food" in the "FOOD" category, as demonstrated in (10) and (11), respectively.

(5) …Stepped back off a curb and hit by a vehicle, bitten by a Snapping Turtle, a bite from a ***poisonous spider***, a nasty case of poisonous oak.

(6) … if it's a ***poisonous mushroom***, that silver will become tarnished.

(7) It hung beneath the ceiling, threateningly thick, a ***poisonous vapor*** that had collected in the apartment.

(8) … a horse dose of heroin and traces of choleric tricemate, which is a ***poisonous chemical*** found in laundry detergent.

(9) But Hercules has a secret weapon, arrows dipped in the ***poisonous venom*** of an earlier conquest, the Hydra.

(10) Gollum leads them into the lair of Shelob, an enormous spider-like creature, who inflicts her ***poisonous bite*** on Frodo.

(11) … as well as the motif of a sword in a stone and of the ***poisonous food*** which deforms two of three brothers.

Table 4. Semantic preference of noun collocates of *venomous*

Categories	Collocates
1. CREATURE	animal, ant, cobra, creature, insect, lizard, pit, rattlesnake, reptile, scorpion, serpent, snake, species, spider, toad, viper
2. ATTACK	attack, bite, snakebite
3. ORGAN	fang, spine, sting
4. ACTION	debate, reply
5. FEELING	hatred

Upon scrutinizing and categorizing the noun collocates of *venomous* into five distinct groups, it was observed that two themes, CREATURE and ATTACK, are common to both *poisonous* and *venomous*, while the remaining three themes seem to be unique to *venomous*: ORGAN, ACTION, and PLACE. In the CREATURE category, a total of 14 nouns—including *animal, ant, cobra, creature, insect, lizard, pit, rattlesnake, reptile, scorpion, serpent, snake, species, spider*, and *toad*—were found to frequently collocate with *venomous*, as demonstrated in (12). Additionally, three nouns—*attack, bite*, and *snakebite*—were categorized under ATTACK, as shown in (13), and three, i.e., *fang, spine*, and *sting*, under "ORGAN," as illustrated in (14). It is worth noting the presence of two noun collocates categorized under ACTION, namely *debate* and *reply*, as depicted in (15). These nouns, which are inanimate and non-object, are rarely associated with *poisonous*. Furthermore, the theme FEELING, with *hatred* as its noun collocate, as illustrated in (16), distinguishes *venomous* from *poisonous*, as the latter typically lacks noun collocates expressing emotions.

(12) Black-headed pythons eat other reptiles, including ***venomous snakes***.

(13) In recent years, the average mortality rate in Texas is about one per year; that hardly classifies death by ***venomous snakebite*** as a high risk, at least in the United States.

(14) The truth is, these bees have an exceptionally ***venomous sting***.

(15) Winkler brilliantly weaves together the dramatic stories of gun rights advocates and gun control lobbyists, providing often unexpected insights into the ***venomous debate*** " Pre-order both items ".

(16) When confronted with resistance to these initiatives, members of today's elite betray the ***venomous hatred*** that lies not far beneath the smiling face of upper-middle-class benevolence.

Based on the findings concerning the usage of *poisonous* and *venomous* across different genres, it is evident that both terms are predominantly utilized in informal written English. Specifically, *poisonous* exhibits the highest occurrence rate in fiction, magazines, and TV/movie subtitles, whereas *venomous* is most prevalent in magazines, fiction, and TV/movie subtitles. Notably, these genres typify informal language usage. In the case of fiction data in the Corpus of Contemporary American English (COCA), it encompasses short stories, plays from literary magazines, children's magazines, popular magazines, first chapters of first edition books from 1990 to the present, and movie scripts. The magazine data comprises a diverse selection of nearly 100 publications, covering various domains such as news, health, home and gardening, women's interests, finance, religion, sports, and more, with a balanced representation over different years. Lastly, the TV/movie series data is sourced from the American segment of the TV and Movies corpora, with subtitles reflecting a degree of informality equivalent to or surpassing actual spoken language, as noted by Davies (2020). The observation that both target synonyms demonstrate similar prevalence across these genres underscores their substantial synonymy, thereby corroborating previous research findings. (e.g., Boontam & Phoocharoensil, 2022; Chaokongjakra, 2023; Kruawong & Phoocharoensil, 2022; Narkprom, 2024; Phoocharoensil, 2020b).

Data on the adjective + noun collocations in which *poisonous* and *venomous* are components demonstrate a close relationship between the two when describing toxic animals or attacks involving poison. They share similar noun collocates such as "creature" (e.g., *animal, snake, species*) and "attack" (e.g., *bite*). It's worth noting that while *poisonous* often collocates with substances like *gas* (e.g., *fume, gas, vapor*) and chemical (e.g., *mercury*), *venomous* typically does not associate with such nouns. Instead, *venomous* tends to collocate with body parts of dangerous animals (e.g., *snakes' fangs, bees' stings, fish's spines*), a semantic category less apparent with *poisonous*. Additionally, *venomous* is found alongside abstract concepts like actions (e.g., *debate, reply*) and feelings (e.g., *hatred*), unlike *poisonous*. This variation in typical collocations highlights differences in usage between these synonyms, which are best explained with authentic corpus data. In the present study, both target synonyms can be differentiated because typical, specific collocations are clearly presented. The efficacy of collocations in serving as a synonym distinguishing criterion is supported by several studies (e.g., Crawford & Csomay, 2016; Chaengchenkit, 2023; Chaokongjakra, 2023; Jirananthiporn, 2018; Kruawong & Phoocharoensil, 2022; Phoocharoensil, 2020a, 2020b, 2021a, 2021b; Narkprom, 2024; Imsa-ard & Phoocharoensil, 2022).

4. Conclusion and Pedagogical Implications

The current investigation illustrates that the synonymous adjectives *poisonous* and *venomous*, which exhibit considerable similarity in meaning according to three learner dictionaries, do not significantly vary in terms of formality, as both are predominantly encountered in informal textual contexts such as fiction and magazines. A more discerning criterion for distinguishing between these nearly synonymous terms lies in their respective collocational patterns. While *poisonous* and *venomous* are frequently associated with toxic animals like snakes and spiders, as well as actions such as *attack*, there are discernible groups of noun collocates that preferentially co-occur with one term over the other. Specifically, words related to gases, chemicals, substances, and food tend to align with *poisonous*, whereas certain

nouns pertaining to actions, emotions, and food are more commonly paired with *venomous*.

ELT practitioners may find the outcomes of this study valuable in several ways. Firstly, drawing from the aforementioned findings, instructors can offer students a clear and persuasive understanding of the nuanced disparities between *poisonous* and *venomous* regarding their distribution across different genres and typical collocations. It is recommended that teachers, leveraging native-speaker English corpus data, underscore to students the fact that very few, if any, absolute synonyms exist, and most English synonyms are considered loose synonyms, meaning they cannot be used interchangeably across all contexts of use. Secondly, educators can employ a similar corpus-based approach to explore other sets of synonyms, integrating common collocations sourced from authentic language corpora to enrich their ELT materials, thereby aiding students in acquiring synonyms through context-based exposure to natural English.

However, it is crucial to consider the limitations of this study. Firstly, the study focused on extracting the top 25 noun collocates based on frequency and MI score, which might not encompass all possible typical collocates. Unlike some previous studies that limited the number of potential typical collocates, this study aimed to categorize 25 collocates based on their semantic preferences, potentially overlooking additional themes that could arise from a broader selection of noun collocates from COCA. Additionally, the criteria used to differentiate the synonyms were limited to two factors: formality, indicated by genre distribution, and common collocations. Future studies can consider distinguishing other synonyms using additional criteria, e.g., connotations or colligations. Finally, the present study examined only the two synonyms, while there seem to be other words whose core meaning is very close to them, e.g., *toxic*. Therefore, future research can also include such synonyms for clearer understanding of the synonyms in this group.

Acknowledgments

Not applicable

Authors' contributions

Not applicable

Funding

Not applicable

Competing interests

Not applicable

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus Linguistics*. Edinburgh: Edinburgh University Press. https://doi.org/10.1515/9780748626908

Barnbrook, G., Mason, O., & Krishnamurthy, R. (2013). *Collocation. applications and implications*. London: Palgrave Macmillan. https://doi.org/10.1057/9781137297242

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Boontam, P., & Phoocharoensil, S. (2022). Broaden your horizons: Distribution and collocational patterns of the English synonyms *expand, widen*, and *broaden*. *The International Journal of Communication and Linguistic Studies*, *20*(1), 107-123. https://doi.org/10.18848/2327-7882/CGP/v20i01/107-123

Carter, R. (2012). *Vocabulary: Applied linguistic perspective*. London: Routledge. https://doi.org/10.4324/9780203124659

Chaengchenkit, R. (2023). A corpus-based study of the synonyms "cease," "halt," and "stop." *LEARN Journal: Language Education and Acquisition Research Network, 16*(1), 473-494.

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), 512-532

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. Retrieved from https://www.youtube.com/watch?v=zW9I9C9rPD4

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 https://doi.org/10.1111/lang.12225

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.

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

Imsa-ard, P., & Phoocharoensil, S. (2022). "A whole new world… wait, Is it a whole, entire, or total world?": The extraction of collocations for the three English synonym discrimination. *3L: Language, Linguistics, Literature. The Southeast Asian Journal of English Language Studies, 28*(2), 67-82. http://doi.org/10.17576/3L-2022-2802-05

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. Retrieved from 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.

Lindquist, H., & Levin, M. (2018). *Corpus linguistics and the description of English*. Edinburgh: Edinburgh University Press.

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

Narkprom, N. (2024). A corpus-based investigation into English synonymous verbs 'restrict' and 'constrain'. *LEARN Journal: Language Education and Acquisition Research Network, 17*(1), 688-714.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

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

Oxford advanced learner's dictionary (2014). Oxford: Oxford University Press.

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. http://doi.org/10.17576/3L-2020-2603-01

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. https://doi.org/10.18848/2327-7882/CGP/v19i01/1-17

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. https://doi.org/10.32601/ejal.911269

Phoocharoensil, S. (2021b). Multiword units and synonymy: Interface between collocations, colligations, and semantic prosody. *GEMA Online® Journal of Language Studies*, *21*(2), 28-45. http://doi.org/10.17576/gema-2021-2102-02

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. https://doi.org/10.1111/lang.12387

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

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), 208-219.

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of trouble with quantitative studies. *Functions and Language*, *2*(1), 23-55. https://doi.org/10.1075/fol.2.1.03stu

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research*. London: Routledge. https://doi.org/10.4324/9781315107769

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.