



# Semantic prosody and collocation: A corpus study of the near-synonyms *persist* and *persevere*

Supakorn Phoocharoensil <sup>a</sup> \* 

<sup>a</sup> *Language Institute Thammasat University, 2 Prachan Road, Bangkok 10200, Thailand*

Received 09 September 2020

| Received in revised form 04 January 2021

| Accepted 25 January 2021

---

## APA Citation:

Phoocharoensil, S. (2021). Semantic prosody and collocation: A corpus study of the near-synonyms *persist* and *persevere*. *Eurasian Journal of Applied Linguistics*, 7(1), 240–258.

Doi: [10.32601/ejal.911269](https://doi.org/10.32601/ejal.911269)

---

## Abstract

Near-synonyms in English often cause considerable confusion among EFL students. This study aims to clarify this confusion through a corpus-based investigation of the target synonymous verbs *persist* and *persevere* with focus on distribution across genres, collocations, and semantic preference/prosody. The results, based on the Corpus of Contemporary American English, reveal that *persist* and *persevere* occur with the highest frequency in academic texts and webpages, respectively. Despite similarities in cognitive meanings, the two synonyms co-select different sets of collocations. An analysis of the semantic prosody suggests that while *persist* tends to co-occur with words or phrases that express negative meanings related to continual unpleasant situations, the co-occurring lexical items around *persevere* denote strong determination and great effort in completing a difficult task. Moreover, although it is stated in a learner dictionary that *persevere* is used only as an intransitive verb, corpus evidence proves that this verb can also be used transitively. It is recommended that corpus data be incorporated in vocabulary instruction, alongside learner dictionaries, the former of which can provide more authentic linguistic information on collocational and grammatical patterns.

© 2021 EJAL & the Authors. Published by *Eurasian Journal of Applied Linguistics (EJAL)*. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

**Keywords:** Semantic prosody; collocation; near-synonymy; corpora; COCA

---

## 1. Introduction

In the context of English language learning and teaching, knowledge of vocabulary is a vital part of mastering English as a second language (ESL) or English as a foreign language (EFL). There is a close relationship between vocabulary, especially measures of vocabulary size, and language proficiency (Schmitt, 2010). One thing that often presents difficulties for learners in their acquisition of L2 English vocabulary is synonymy, i.e., a relationship of sameness of meaning that may exist between two words (Jackson & Amvela, 2007). Several learners assume that they can replace a word with its synonym without any change in meaning and structure. Such a

---

\* Corresponding author. Tel.: +662-6133101

E-mail address: [supakorn.p@litu.tu.ac.th](mailto:supakorn.p@litu.tu.ac.th)

<http://dx.doi.org/10.32601/ejal.911269>

substitution of near-synonym can cause ungrammaticality or unnatural production in L2 English (Thornbury, 2002) since, in actuality, there appear no words in English that can perfectly occur in exactly the same context of use (Jackson & Amvela, 2007).

A number of studies have shown that near-synonyms differ in many aspects, including degree of formality (e.g., Jirananthiporn, 2018; Petcharat & Phoocharoensil, 2017), connotations (e.g., Partington, 1998; Stubbs, 1995), collocations (e.g., Crawford and Csomay, 2016; Jirananthiporn, 2018; Phoocharoensil, 2010, 2020), semantic prosody (e.g., Nelson, 2006; Partington, 1998, 2004; Selmistraitis, 2020), and grammatical patterns (e.g., Phoocharoensil, 2010).

In this study, the pair of near-synonyms *persist* and *persevere* was analyzed in terms of genres, collocations, and semantic preference/prosody. The main purpose is to make a clear distinction between both verb synonyms through data drawn from a massive corpus representing American English, i.e., the Corpus of Contemporary American English (COCA). The meanings and usages of the target synonyms from a learner dictionary were compared and contrasted before COCA data was taken into account for a clearer image of the way both words occur in authentic English.

As the focus of the analysis was mainly on collocations and semantic preference/prosody to which *persist* and *persevere* are attached, these concepts will be explored in the following section.

## 2. The Concepts of Collocation, Semantic Preference, and Semantic Prosody

### 2.1. Collocations

It is widely accepted that the concept of ‘collocation’ plays a prominent role in English Language Teaching (ELT) and Applied Linguistics. It refers to a multiword unit or formulaic language that occurs with very high frequency and semantically or syntactically forms a meaningful unit (Moon, 2010). Firth (1957) was probably the first to highlight the importance of lexical partnerships represented by collocations, pointing out that word meaning does not only lie in the word itself but is also dependent on the way it is combined with the words around them. Scholars define the term ‘collocation’ in different but related ways. Sinclair (1991) defined collocation as “items that occur physically together or have stronger chances of being mentioned together” (p. 170). Similarly, Lewis (1997) defines collocation as “the readily observable phenomenon whereby certain words co-occur in natural text with greater than random frequency” (p.8). Collocation occurs in statistically significant manners (Lewis, 2000). An example of common English collocations is the adjective + noun collocation *strong tea* rather than *\*powerful tea*.

English learners require sufficient collocational knowledge in order for them to “speak and write English in a more natural and accurate way” (McCarthy & O’Dell, 2005, p. 4). Making use of word combinations that are rare in natural English or very low in the degree of collocability can result in non-standard English (Thornbury, 2002;

Szudarski, 2018). For instance, despite sharing the meaning of ‘having unpleasant smells or tastes’ in describing food that is not fresh, the adjective synonyms *rancid* and *sour* typically co-occur with various kinds of food. It is much more common to say *rancid butter* than *\*sour butter*, and, in a similar vein, *sour milk*, as opposed to *\*rancid milk*, forms a strong, natural collocation (Crawford & Csomay, 2016, p. 40).

Kuiper, Columbus & Schmitt (2009) discovered that formulaic language, including collocations, offer processing efficiency, as single memorized units made up of a string of words are processed faster than word sequences generated creatively. Furthermore, non-native speakers of English were found to read and understand formulaic sequences more quickly than non-formulaic units (Conklin & Schmitt, 2008). To summarize, knowledge of collocations appears to facilitate the thinking process, which leads to fluent communication. Due to its significance in second language acquisition and instruction, collocations should be incorporated in vocabulary lessons. That is, “...language teachers need to be equipped with a proper understanding of the importance of collocation in language acquisition, and trained in its effective application in the classroom” (Barnbrook, Mason, & Krishnamurthy, 2013, p. 121)

With regard to near-synonym investigation, past studies arrived at similar conclusions in that although some words are similar in cognitive or denotational meanings, they could be different in collocational or prosodic behavior. It has been proven that “...synonymous words are not collocationally interchangeable” (Xiao & McEnery, 2006, p. 108). Language learners are required to have a clear understanding of specific circumstances in which a word can be replaced by another with similar meaning, and, more importantly, the context in which such a substitution reduces the degree of acceptability or naturalness in L2 English collocations (Barnbrook, Mason, & Krishnamurthy, 2013).

Partington (1998) pinpointed the critical role of collocations and patterns in determining the subtle differences between the near-synonyms *sheer*, *pure*, and *complete*. Even though the three adjectives describe a state or situation that is complete and does not involve or is not mixed with anything else, their collocational behavior differs. The typical phraseology of *sheer* is *the sheer + magnitude word + of + noun phrase*, as in *the sheer weight / number(s) / magnitude / mass / scale / quantity / multiplicity / size / tonnage of*. In addition, *sheer* also appears with noun collocates that express ‘persistence’, such as *obstinacy, persistence, insistence, resolve, hard work, irreversibility*, etc. Despite a degree of collocational overlap with *sheer*, *pure* can exclusively modify material substances, as in *pure wool*, and also has a religious-moral semantic content, as in *a pure doctrine of religion, pure faith, and pure goodness*. As another synonym of *sheer*, *complete* behaves like *sheer* in that it is sometimes used in the pattern *the complete + noun / noun phrase*. However, the noun occurring immediately after *complete* is not a multitude or force word. Examples of noun collocates of *complete* are *absence, withdrawal, transfer, destruction, abandonment, removal, disintegration*. Likewise, the near-synonym *absolute* has a lexical sense of ‘unchallenged’ or ‘unquestioned’, as in *absolute command* and *absolute limits of power*, a sense which is not existent in *sheer*. Partington concluded that

collocation is a key factor determining synonym differences since “the substitution of one [synonym] by another may not change the meaning of the phrase in any evident way, but one item may well be more appropriate than another...” (p. 39).

Collocation was also the major criterion used in distinguishing two adjective synonyms *equal* and *identical* in a study by Crawford and Csomay (2016). Generally, it was shown that both synonyms are interchangeable in certain contexts; however, in an analysis of collocations derived from COCA, different collocational behavior was revealed in that there is more likelihood of abstract concepts such as *opportunities*, *rights*, and *protection* accompanying *equal*, and concrete nouns such as *twins*, *copies*, and *houses* co-occurring with *identical*. Such an observation is only possible by referencing large amounts of texts through corpus data, as it allows linguists and language researchers to access language use in authentic contexts.

Dedicated to the interface between genres and collocations in differentiating the synonyms *problem* from *trouble*, Jirananthiporn’s (2018) study demonstrated the key differences between both synonyms using data from COCA. Having thoroughly examined the occurrences of the two words across five genres, i.e. academic texts, popular magazines, newspapers, fiction, and spoken language, the researcher reported on the frequencies of *problem* far outnumbering *trouble* in all the genres available in COCA. In terms of collocations, *problem* often co-occurs with verbs associated with written discourse, many of which are multi-syllable lexical verbs, e.g. *alleviate*, *eliminate*, *investigate*, *recognize*, *exacerbate*, while *trouble* is more likely to be combined with phrasal verbs, e.g. *head for*, *run into*, or one-syllable verbs, e.g. *have*, *give*, *start*, which explicitly characterizes spoken English.

A recent study distinguishing near-synonyms based on the genres and collocations provided by the latest version of COCA (2020) is Phoocharoensil (2020). Three noun synonyms, i.e. *consequence*, *result*, and *outcome*, were focused in the study. It was revealed that the three target words occur more frequently in academic texts, which means they are mainly associated with formal language. Despite sharing some denotative meaning, the analysis of typical collocations clearly showed that *consequence* is very common in adversative contexts; some examples of its frequent adjective collocates are *catastrophic*, *dire*, *disastrous*, *inevitable*, *unfortunate*, while *result* is closely connected with research methodology, as can be seen in its common verb collocates such as *corroborate*, *generalize*, and *replicate*. *Outcome*, unlike the other two synonymous nouns, apparently collocates with words with a wide range of semantic properties, such as *academic*, *beneficial*, *final*, *predictable*, *long-term* etc.

## 2.2. Semantic Preference vs. Semantic Prosody

Closely related to the concept of collocation are ‘semantic preference’ and ‘semantic prosody’, both of which are concerned with co-selections made by speakers and writers. Semantic preference is defined as “the tendency for lexical items to be restricted to identifiable semantic fields” (Cheng, 2012, p. 114). It is also referred to as “the relation between an individual word and a set of semantic categories”

(Flowerdew, 2012, p. 323). In other words, semantic preference is viewed as a semantic environment in which vocabulary items commonly occur.

Some observations can be made on the nature of semantic preference (Flowerdew, 2012). First, semantic preference for a specific domain is usually attached to the syntactic environment where it appears. Word meaning is bound to particular grammatical constructions. To illustrate, Partington (2004) reported on his discovery of the semantic preference of the verb *cause*. When *cause* precedes a single object, the object is frequently an illness, e.g. *cancer*, *heart disease*, but when double objects follow *cause*, the second object often expresses unpleasant feeling or emotion, e.g. the noun *convenience* in *causes them inconvenience*.

The second observation pertains to the varying degree of semantic preference according to the corpus type. This means the type of corpora on which an investigation is based is a major determinant of the semantic preference of a lexical item. In support of this, Nelson (2006) found that the semantic preference of the adjective *global* varies from one kind of corpus to another. To be more precise, in a 1-million-word business English corpus (BEC), *global* is very rich in semantic sets such as *global products* and *global economic indicators*. In contrast, only two semantic sets in which *global* is involved are found in a general corpus like the British National Corpus (BNC). These two sets are related to climate, as in *global warming*, and people, as in *global viewer*.

In making a distinction between synonymous words, semantic preference is usually analyzed together with collocations (Szudarski, 2018). As mentioned earlier, Partington (1998) made an effort to distinguish the adjective *sheer* from its near-synonyms *pure*, *complete*, and *absolute*. As exploring the possible collocates of *sheer* through concordance lines from academic and newspaper corpora, Partington (1998) categorized these noun collocates based on their semantic properties into four semantic categories:

- (a.) noun collocates expressing magnitude, weight, or volume, e.g. *the sheer volume of reliable information*, *the sheer weight of noise*, *the sheer scale of prizes* (pp. 34-35)
- (b.) noun collocates expressing force, strength, or energy, e.g. *the sheer physical energy*, *the sheer force of an earthquake*, *the sheer power* (p. 35)
- (c.) noun collocates expressing persistence, e.g. *sheer persistence*, *the sheer obstinacy*, *his sheer indomitability* (p. 36)
- (d.) noun collocates expressing strong emotion, e.g. *its sheer inspiration*, *the sheer joy in life*, *sheer panic*, *sheer exhilaration* (p. 37)

It can be seen that the adjective synonyms *pure*, *complete*, and *absolute* occur in their own particular set of phraseological patterns, which implies that none of them

can be interchangeably used with *sheer* in spite of their similarities in meaning. Partington (1998) maintained “every lexical item in the language has its own individual and unique pattern of behavior” (p. 46). Thus, English learners who consult a thesaurus for lists of synonyms may face difficulties in correctly using words for specific contexts.

In relation to semantic preference, semantic prosody refers to “a consistent aura of meaning with which a word is imbued by its collocates” (Louw, 1993, p. 157). According to Sinclair (2004), semantic prosody integrates a word with its collocates. This notion is considered particularly important to the study of phraseology as the concept concerns “the spreading of connotational colouring beyond single word boundaries” (Partington, 1998, p. 68).

Hunston (2002) clarified the features of semantic prosody as follows. First, the semantic prosody of a lexical item deals with meaning derived from consideration of the entire bundle or string of words rather than a single word. Second, semantic prosody reflects the connotation of a word in context. A favorable or unfavorable connotation is not always clearly imparted by a single word alone but instead is represented by that particular word in association with its collocates (Partington, 1998). The verb *commit* is an example of a word having negative connotation expressed by a neighboring word, as in *commit deception*, *commit acts of violence*, *commit serious crimes*, and *commit suicide* (Partington, 1998, p. 67). Observation of the unfavorable connotation associated with *commit* is only possible on the condition that a large number of instances of a word or phrase are accessible as semantic prosody “relies on the typical use of a word or phrase” (Hunston, 2002, p. 142). This is the third key feature of semantic prosody. In accordance with Hunston (2002), Szudarski (2018) suggested that the semantic prosody of a lexical item “comes to light only when we are presented with large amounts of authentic data, which allows us to explore different patterns and configurations of lexical and lexico-grammatical units” (p. 86).

A well-cited corpus-based study on semantic prosodies was conducted by Stubbs (1995), focusing on the verbs *cause* and *provide*. Based on 40,000 examples from 120 million words from the Cobuild Corpus, the corpus analysis indicates that *cause* has a high tendency to co-occur with nouns expressing adversity, e.g. *anxiety*, *problem*, *confusion*, *damage*, *death*, *delay*, *harm*, *trouble*, etc., whereas *provide* typically collocates with words of a positive sense, e.g. *care*, *food*, *growth*, *help*, *money*, *progress*, *joy*, etc. Hunston (2002) notes that access to the connotation or subtle meaning embedded in a word is not permissible from the native speaker’s conscious knowledge. Partington’s (2004) investigation of the phrasal verb *set in* supports this claim in that *set in*, which generally means ‘begin’, is often found to collocate with negative words. In agreement with Partington (2004), Lindquist and Levin’s (2018) exploration of *set in* in COCA shows that the subject noun collocates are largely indicative of negative meanings, e.g. *vengeful habits*, *panic*, *economic depression*, *lawlessness*, *authoritarianism*, *the cold war*, *the psychological toll of unemployment*, *arthritis*, etc. With corpus-informed data exhibiting semantic association between a lexical item and

its surrounding words, it can be said that semantic prosody carries attitudinal or evaluative meaning “which results from a word’s co-occurrence with specific collocations” (Szudarski, 2018, p. 86).

Similar to the behavior of semantic preference, the semantic prosody of a word can be global or local depending on the corpus from which data are obtained. A global prosody is reflected in language in general, whereas a local prosody is concerned with the linguistic environment is a specific genre or discipline (Flowerdew, 2012). To illustrate the difference between both kinds of prosodies, Partington (2004) studied *lavish* in two different genres and found that *lavish* is frequently combined with noun collocates indicating disapproval in newspapers. However, this particular prosody does not exist in normal British conversation.

Selmistraitis (2020), in addition, explored the semantic preference/prosody of three pairs of adjective synonyms, namely *succinct & concise*, *coherent & cohesive*, and *precise & accurate* in the academic texts of Corpus of Contemporary American English (COCA). Despite the fact that the target pairs of synonyms are similar in meaning, they differ in semantic preference and prosody. In particular, *concise* seems to co-occur particularly with nouns referring to books, e.g. *dictionary* and *encyclopedia*, whereas *succinct* does not. The common collocates of *cohesive* are often collective nouns denoting a number of people, e.g. *group*, *unit*, *team*, *community*, those of *coherent* denote a plan of actions for particular purpose, a set of ideas or description, e.g. *policy*, *strategy*, *framework*, *theory*, *vision*. The word *precise* is normally combined with words expressing temporality and locality, e.g. *moment*, *date*, *timing*, *location*, *contours*, while *accurate* shows preference for abstract nouns denoting ‘proving’, ‘identification’, ‘discovering’. Interestingly, the target synonyms are associated with neutral semantic prosody, borne out by the fact that they usually collocate with neutral vocabulary. This may be due to the fact that the genre being examined was the academic English, in which neutral semantic prosody is the norm.

Furthermore, colligation, i.e. the grammatical environment in which a word usually appears, also plays a significant role in determining the semantic prosody (Flowerdew, 2012). For instance, it was discovered that the structure of *the + adjective* in referring to an entire class of people often contains adjectives expressing negative meanings, e.g. *the elderly*, *the unemployed*, *the sick* (Sinclair, 2004). In addition to colligation, transitivity sometimes has a correlation with the semantic prosody of verbs. Louw (1993) showed that when *build up* is used as a transitive verb after a human subject, the semantic prosody often appears to be positive, with this multiword verb preceding objects such as *organizations*, *understanding*, *trust*, etc. In contrast, negative prosody results if *build up* occurs as an intransitive verb with subjects such as *cholesterol*, *toxins*, *dirt*, etc.

In this paper, I will look at two near-synonyms *persist* and *persevere*, which have very similar cognitive meanings, paying particular attention to their distribution across genres, and collocational behavior in relation to semantic preference and prosody. It is predicted that both words differ in the noun collocates with which they

frequently co-occur, and the difference in collocational patterns will account for their differing prosodic behavior.

### 3. Methods

This study began with consulting a learner dictionary, i.e., Longman Dictionary of Contemporary English, 6<sup>th</sup> Edition (LDOCE), for the definitions, sentence examples, and usage notes of the target synonyms. Next, the Corpus of Contemporary American English (COCA) was examined for frequencies of the two synonyms across genres and noun collocates. As one of the largest and most well-known English corpora for ELT researchers and practitioners, the latest version of COCA (Davies, 2020) was launched in March 2020 and encompasses texts of eight different genres, consisting of five conventional genres, i.e., spoken, fiction, popular magazines, newspapers, and academic texts, and three new genres, i.e., TV and movie subtitles, blogs, and webpages. COCA contains over one billion words, with approximately 20 million words being included annually from 1990-2019.

In the following step, nouns typically collocating with *persist* and *persevere* were listed based on the collocational strength determined by the Mutual Information (MI) value (or score). The Mutual Information value is calculated through statistics which confirm whether two words co-occur randomly or have a strong association in terms of collocation. However, sole dependence upon the MI score is not the best means of collocation extraction due to the fact that collocations with very high MI scores may be extremely low in total frequency in a corpus (Schmitt, 2010). Thus, to avoid listing noun collocates that are rare occurrences despite very high MI value, the strength of collocation was assessed by the frequency of noun collocates in conjunction with the MI value (Cheng, 2012; Gablasova, Brezina & McEnery, 2017). The noun collocates that are in the top-20 frequency list shown in COCA and MI score of which is  $\geq 3$ , which is the significance value for collocational association, were included (Cheng, 2012).

After the extraction of noun collocates, the nouns with which *persist* and *persevere* frequently co-occur were categorized based on the semantic preference, or similarities in meaning. In analyzing the semantic prosody, which is regarded as a subset of semantic preference (Flowerdew, 2012) of the two near-synonyms, 100 concordance lines of each target word randomized from COCA were thoroughly scrutinized for connotations specific to each individual verb synonym.

### 4. Results and Discussion

The definitions of *persist* and *persevere* given in LDOCE (2014, p. 1351) are as follows:

**persist**



1. [intransitive, transitive] to continue to do something, although this is difficult, or other people oppose it

***persist in (doing) something***

e.g. He persisted in his refusal to admit responsibility.

***persist with***

e.g. She persisted with her studies in spite of financial problems.

‘I don’t think it’s right,’ John persisted.

2. [intransitive] if something bad persists, it continues to exist or happen

e.g. If the pain persists, you must see a doctor.

**persevere** [intransitive] to continue trying to do something in a very determined way in spite of difficulties – use this to show approval

e.g. It can be tricky at first, but persevere.

**persevere with**

e.g. He persevered with his task until he had succeeded in collecting an armful of firewood.

**persevere in (doing) something**

e.g. She had persevered in her claim for compensation.

As can be seen in the above definitions, *persist* and *persevere* can be treated as near-synonyms as they share the general cognitive and denotational meaning of ‘continue doing something in a difficult situation’. Aside from the core meaning, they are usually used in the same colligational patterns, namely *persist/persevere in (doing) something* and *persist/persevere with + noun*. Based on LDOCE, both synonyms differ in transitivity since *persist* can be used as a transitive verb or an intransitive verb, while *persevere* only occurs as an intransitive verb. However, it is worth adding that there exists no example of *persist* used transitively in the hardcopy or online version of LDOCE, the latter of which includes considerably more sentence examples. It could be assumed that the main usage of *persist* in terms of transitivity is very similar to that of *persevere*. LDOCE also implies some semantic prosody of *persist* when appearing in a negative environment, as indicated in the second meaning, which is also semantically related to the first in continuity. Unlike high-frequency words for which some common collocations are given in LDOCE, no collocational information of the two verbs is available in the dictionary.

The findings in regards to distribution across genres in COCA will now be presented and discussed. In the following step, nouns typically collocating with *persist* and *persevere* will also be explored.

Table 1. Distribution of the synonyms *persist* and *persevere* across eight genres according to frequency

<i>Persist</i>			<i>Persevere</i>		
Genre	Frequency	Per million	Genre	Frequency	Per million
academic texts	1,775	14.82	webpages	209	1.68
magazines	859	6.81	blogs	167	1.30
webpages	764	6.15	academic texts	158	1.32
blogs	593	4.61	magazines	129	1.02
newspapers	490	4.02	newspapers	110	0.90
spoken	239	1.89	spoken	82	0.65
fiction	227	1.92	TV and movie subtitles	54	0.42
TV and movie subtitles	91	0.71	fiction	61	0.52
<b>Total</b>	<b>5,038</b>		<b>Total</b>	<b>970</b>	

In Table 1, the total frequency of *persist* (5,038 tokens) is far higher than that of *persevere* (970 tokens), and occurs with more frequency than *persevere* in each genre in COCA. In terms of formality, the occurrences of *persist* are most frequent in academic texts (1,775 tokens), while the highest frequency of *persevere* is found in webpages (209 tokens), which shows that *persist* is associated with a higher level of formality than *persevere*. However, if we consider the top-4 genres in which both synonyms appear, the two verbs appear to be common in written English as they are more frequently used in magazines, webpages, blogs, and academic texts, i.e., written genres. The corpus data also indicates that *persist* and *persevere* do not prevail in colloquial English due to their relatively lower numbers of tokens in spoken language, fiction, and TV and movie subtitles, i.e., informal styles or contexts. Their low frequency in spoken English genres confirms their high degree of formality, with *persevere* being slightly less formal due to its frequency in academic texts in the third rank compared to *persist*. The finding that *persist* is associated with academic language is consistent with Coxhead’s (2000) Academic Word List (AWL), in which *persist* is included.

Table 2. Noun collocates of *persist* and *persevere* from COCA

Rank	<i>Persist</i>			<i>Persevere</i>		
	Noun collocate	Frequency	MI-Scores	Noun collocate	Frequency	MI-Scores
1	symptom	351	5.01	faith	22	4.00
2	rumor	96	5.36	prayer	6	3.15
3	pattern	92	3.49	Christ	6	8.69
4	gap	63	4.06	believer	3	3.69
5	trend	58	3.50	devotee	3	6.47
6	myth	49	4.40	Jesus	3	7.56

7	drought	45	5.29	mentor	2	3.08
8	inequality	39	5.30	filmmaker	2	3.64
9	tension	32	3.36	Nazi	2	3.95
10	poverty	31	3.14	China	2	4.00
11	stereotype	27	4.52	lad	2	4.32
12	racism	25	3.77	martyr	2	4.66
13	disparity	25	5.13	America	2	5.71
14	unemployment	24	3.21	robin	2	5.91
15	discrimination	24	3.49			
16	confusion	19	3.23			
17	intent	17	3.25			
18	uncertainty	16	3.16			
19	shortage	16	3.61			
20	suspicion	15	3.17			

In terms of collocations, both *persist* and *persevere* co-occur with different groups of nouns. In Table 2, there exist 20 frequent noun collocates of *persist* and only 14 of *persevere* based on frequency and MI scores ( $\geq 3$ ). It should be cautioned here that the collocational information provided by COCA needs to be dealt with carefully for two major reasons. First, although several words are listed in COCA as noun collocates of the search words, not all of the nouns function as the grammatical subject of *persevere*; some of the nouns may appear in the position near this target verb without having any subject-verb relationship, and thus were excluded from the analysis. In brief, the subject noun collocates of *persevere* are usually human, with the exception of *faith*. Moreover, some pronouns, e.g. *I*, *we*, *they*, often occupy the subject position of *persevere*, but a combination made up of such a particular pronoun, as in *we persevered*, does not constitute a strong collocation since these pronouns, which are widely used with other words in a variety of contexts, are not specifically or mutually used with *persevere* in a statistically significant way. Such pronouns were therefore not included in the present-study analysis either. Second, as for *persist*, some nouns occurring with high frequency, e.g., *problem*, *pain*, are not presented in Table 2 because their MI score is lower than 3, i.e., the level of significance.

Having made a comparison between the noun collocates of *persist* and *persevere*, it was discovered that *persevere* co-selects only 14 noun collocates, whereas *persist* has more than 20 subject nouns with which it can collocate. That *persist* can be combined with a wider range of noun collocates than the other near-synonym probably results from its far higher frequency in English, instantiated by the COCA data. In other words, there is more chance for *persist* to appear with many more collocates than a synonym like *persevere* whose frequency is much lower. Furthermore, the two

synonymous verbs do not share any noun collocates, which means they have different collocational behaviors despite the meaning overlap. At this point, it can be seen that the number of noun collocates accompanying *persist* and *persevere* helps distinguish both synonyms to a certain degree in terms of the variety of collocations.

The collocating nouns listed in Table 2 were then grouped according to their semantic preference. Two principal semantic sets cover the noun collocates of *persevere*. The first set is restricted to the religious context of Christianity, including words such as *faith*, *prayer*, *Christ*, *believer*, *devotee*, *Jesus*, *martyr*. Despite the fact that most of the subject nouns co-occurring with *persevere* are human, the non-human noun *faith* in the combination *faith perseveres* is common in the Christian context. The second semantic category having fewer members includes words referring to powerful countries, e.g., *China*, *America*, or an ideology, e.g., *Nazi*.

An analysis of noun collocates of *persist* revealed further semantic sets. The largest set contains words bound up with sociocultural problems, e.g., *discrimination*, *disparity*, *gap*, *inequality*, *poverty*, *racism*, *stereotype*, and *unemployment*. Another semantic set includes words germane to false or unconfirmed beliefs, e.g., *myth*, *rumor*, *suspicion*, and *uncertainty*. In addition, two words are associated with health problems, i.e., *symptom* and *tension*, while two other nouns relate to disasters, i.e., *drought* and *shortage*. Considering all of the top-20 noun collocates of *persist* confirms the information from LDOCE in that there is a close connection between the verb and the negative meanings it conveys.

Interestingly, there are some noun collocates of *persist* that are seemingly neutral, e.g. *pattern* and *trend*. However, with closer scrutiny into the concordance lines from COCA, it can be seen that these nouns denote negative or unpleasant senses, as shown in the sample concordance lines below:

- 1 Star was just a way to save money. The same **pattern persisted** in Last Crusade, and its why that film is the weakest in
- 2 first three months of this year, but the basic **pattern persisted**: very slow growth, with pockets of outright contraction.
- 3 were at higher risk for poorer disease control and that this **pattern persisted** 2 and 3 years after diagnosis. These findings
- 4 This **pattern persisted** after the two-week delay, however the difference was not statistically significant
5. He may have distinguished between skin color and race, but a basic **pattern persisted** in the racial prerequisite cases: the ascription of darkness increased the chances of ineligibility.

The adversative meaning conveyed by the combination *pattern persisted* becomes noticeable when their occurrences in context are taken into consideration. In other words, a closer investigation into this collocation in relation to neighboring words has revealed the speakers' or writers' unfavorable connotation owing to the negative environment in which *pattern persisted* occurs, as can be seen by surrounding lexical

items expressing adversity, e.g., *weakest, slow growth, higher risk, poorer disease control, two-week delay, the chances of ineligibility*.

This explanation is also true for the noun *trend*, which generally appears to be neutral, but analysis of the semantic prosody of *trend persists* yields a more vivid picture of the negative sense with which this combination is associated. Consider the following concordance examples of *trend persists* from COCA:

1 a mortality rate over 5% the highest in the world, if this **trend persists**, millions of Argentinians will die of viral influenza pneumonia. 5) July 2009

2 expect it to come ashore as a significant hurricane, even if the weakening **trend persists**. What do you expect for Port

3 diet and inactivity, an increase of 33 percent since 1990. If the **trend persists**, obesity will soon replace smoking as the number one cause of preventable death in

4 but even this has barely arrested the declining attendance figures. If the current **trend persists** it is likely that within two years ticket receipts will no longer support the costs

5 been rising against most Western European currencies and the Japanese yen. If this **trend persists**, the 1991 revenues of U.S.-based multinationals will be penalized by currency conversion

6 spending less and talking about saving more and paying down debt. If the **trend persists**, interest rates will spiral down and stay there.

Focusing on the semantic prosody of *trend* as a collocate of *persist* in the above concordance lines sheds light on the negative or unfavorable connotation attached to both words. That is, *trend persists* clearly occurs in the same environment as other words or phrases connected with something unpleasant or undesirable, e.g., *a mortality rate, die of viral influenza pneumonia, hurricane, weakening, inactivity, obesity, death, declining attendance figures, penalized, debt, spiral down*.

To acquire an in-depth understanding of the connotation that *persist* and *persevere* express, 100 randomly selected concordance lines of each near-synonym were analyzed for the semantic prosody, with an emphasis on not merely the adjacent collocates but also other words preceding and following the target words in that context. The concordance lines reveal that although both verb synonyms are close in cognitive meaning, *persist* basically conveys negative meanings, while *persevere* does not. Below are some instances of concordance lines of *persist* obtained from COCA.

1 e-mails that would have resulted in my name being compromised, and seemed to **persist** in these threats after I suggested the possible consequences.

2 I don't wish to **persist** arguing on and on, getting nowhere with you guys.

3 even harder to blame George W. Bush for the stagnant economy that is likely to **persist**.

4 how it's put together, how it functions, and what causes societies to **persist** in their oppressive structure?

5 people behave irrationally, so too does the market, and unfortunately these conditions can **persist** for long period of times.

6 These treacherous illusions **persist** despite government regulation. 7

7 If still cynicism and hatred among others **persist** then let them be.

8 New York Governor Cuomo predicts that state-wide gas shortages will persist for several more days,

9 The potential for the disastrous rise of misplaced power exists and will **persist**.

10 Active residues of pharmaceuticals have been detected in surface water, and they may **persist** in the environment for long periods of time.

11 pain in the area of the hip joint, the result may be that symptoms **persist** and can often give rise to further surgery on the spine

12 The exact length of the operation will depend on how long power outages **persist** and how many homes are affected.

13 It is an act of crass stupidity to **persist** with a losing tactic, and to stand behind strategy of spin and exaggeration

14 are poor competitors or women who are strong competitors, the gender-based preferences for competition **persist**.

15 and hence, our perturbation to the atmospheric concentrations will **persist** for thousands of years.

The data suggests that it is quite common for *persist* to be used in negative environment as the semantic prosody analysis shows its strong association with adversity, which is in line with LDOCE. Subject noun collocates in the lines above demonstrate words with negative meanings, e.g., *the stagnant economy, unfortunately these conditions, active residues of pharmaceuticals, symptoms, power outages, the gender-based preferences for competition, and our perturbation to the atmospheric concentrations*. Apart from the noun collocates preceding *persist*, there are other surrounding words and phrases signifying adversity, e.g., *these threats, arguing on and on, harder to blame, their oppressive structure, behave irrationally, unfortunately, the disastrous rise of misplaced power, pain, an act of crass stupidity, etc.*

An example of *persist* used as a transitive verb as stated in LDOCE may also be found although no sentence example of grammatical usage is available in the dictionary. The data apparently indicates that *persist* is normally used intransitively, followed by the prepositional phrases beginning with *in* or *with* or ending a sentence. However, in example 2 above, *persist arguing*, in which *arguing* acts as an object of the verb, proves that it can occur as a transitive verb as well. This corpus-informed evidence confirms the benefits of using language corpora in tandem with dictionaries in English language teaching and learning (Partington, 1998; Phoocharoensil, 2010).

In comparison to *persist*, the semantic prosody of *persevere* as shown in the following concordance lines were examined.

1 better for having completed the hike and thankful to my friends for forcing me to **persevere**. Good luck at the marathon!

2 As is that thing that tells us we must **persevere** with all our commitments with 100% effort even when we have taken on

3 the book before you as well as from those who disbelieved, but if you **persevere** patiently, and guard against evil

4 he feels very far from God -- we too must be saints and **persevere** in prayer in the midst of dryness and spiritual aridity!

5 the miracle of 'normal development' to the truth of what it means to **persevere**, to keep trying and failing many times, for weeks or months

6 Please give me the strength and faith to **persevere** so that I may bring glory to your name while withstanding the burden of

7 Larry Page is another example of a businessperson who can **persevere** any challenge. Larry and his company have faced much criticism

8 by saying in my experience the glittering prizes in life come more to those who **persevere** despite setback and disappointment

9 He had courage, foresight and steely self determination to **persevere**, without hesitation.

10 by sacrificing to put them in a Catholic school. Data proves they **persevere** in the faith at higher rates, pray better, are more faithful to Sunday

11 It's hard, but if you're interested and **persevere**, you can be successful.

12 the battles you must face and the way in which you **persevere**. I ask that you trust in God

13 Sure, many of the most dedicated activists **persevere**, but the rank and file who made up those teeming crowds don't like

14 unless enough people are informed, organized, determined, and willing to **persevere**. It won't be won unless some effective leaders step forward to rally

15 I watched as Kristy discovered how to **persevere** through difficulties, obstacles, and opposition with commitment, hard work, and determination.

A detailed analysis of the semantic prosody reflected by *persevere* together with other words in the same environment merits some observations. First, up to 62 out of 100 concordance lines are bound to the context of Christianity, 5 of which are exemplified in 3, 4, 6, 10, and 12, shown above. From these instances, the lexical items or bundles associated with Christianity are *God, saints, prayer, disbelieved, guard against evil, in the midst of dryness and spiritual aridity, Please give me the strength and faith, a Catholic school, are more faithful to, I ask that you trust in God*. Second, *persevere* usually co-occurs with words or phrases indicating hardship or difficulties that does encounter, e.g., *hard, hike, setback, marathon, obstacles, difficulties, disappointment, hard work, opposition with commitment, withstanding the burden, battles you must face, have faced much criticism, It won't be won unless*.

Furthermore, vocabulary items relating to strong determination, patience, and endeavors to surmount obstacles or challenges were also discovered in the context of *persist*, e.g. *courage, effort, foresight, determined, organized, willing, commitment(s), sacrificing, determination, steely self determination, without hesitation, the most dedicated activists, keep trying and failing many times*.

A further observation from the concordance lines concerns *persevere* being used as a transitive verb despite the fact that it is labeled as an intransitive verb in LDOCE. This supports Phoocharoensil (2010)'s claim that corpus-informed data often outweighs information from learner dictionaries since language corpora seem to provide more authentic usage with respect to grammatical and collocational patterns, to mention but a few.

The findings of the present study show that synonymous words can be differentiated through consideration of collocations, alongside the semantic preference/prosody with which they are frequently combined. Such dependence on both the collocational patterns and the prosodic behavior of *persist* and *persevere* allows for greater possibilities of discriminating the two near-synonyms effectively, which are consistent with previous studies in that near-synonyms, despite some overlap in core meaning, tend to co-select particularly different sets of synonyms (e.g. Crawford & Csomay, 2016; Jirananthiporn, 2018; Partington, 1998, 2004; Phoocharoensil, 2020; Selmistraitis, 2020). In this study, *persist* is more like to appear with lexical items denoting negative meanings, whereas *persevere* strongly collocates with Christianity-oriented vocabulary and lexical items related to difficulties and determination. To summarize, while some other pairs of synonyms may, to a certain degree, have some collocations in common, the two verb synonyms substantially differ in collocational behavior, and this explains why they rarely share semantic preference or prosody. This confirms the slim chance of collocational interchangeability between near-synonyms (Xiao and McEnery, 2006).

## 5. Limitations and Recommendations for Further Studies

The present study is not without limitations. Only one type of collocation, i.e., noun + verb, is the main focus of the study. Apart from noun collocates extracted from the corpus, other possible collocates of *persist* and *persevere*, e.g., adverbs, in relation to the semantic prosody attached to these collocations, should be investigated in future research. This may provide a clearer picture of the subtle usage of the two synonyms to a greater extent. Another limitation deals with the statistics. As the current study relies on the total frequency and the MI score, the extracted collocates can be different if other statistical tests, e.g., T-score, Z-score, log-likelihood, are applied in further studies to measure the strength of collocational association (Phoocharoensil, 2020).

## 6. Conclusion and Pedagogical Implications



A challenge for many EFL learners in L2 vocabulary acquisition is the correct usage of near-synonyms. As shown in this study, learner dictionaries alone may not demonstrate comprehensive usage of each synonym. Fortunately, with the existence of language corpora, learners are allowed to acquire a wide range of linguistic information of words they are interested in, such as distribution across genres, collocations, and semantic preference/prosody. In terms of distribution, *persist* occurs most frequently in academic texts, while the highest number of occurrences of *persevere* is in webpages. It has also been demonstrated in the current study that although they are similar in cognitive meaning, the near-synonyms *persist* and *persevere* are largely accompanied by different noun collocates that belong to different semantic categories. Analysis of not only the noun collocates but also other surrounding lexical items signifying the semantic prosody reveals the genuine meaning or connotation to which the synonyms are attached. In particular, while *persist* appears with vocabulary items referring to social and cultural problems, untrue or unconfirmed beliefs, unhealthy conditions, and disasters, the typical collocates of and the words or phrases in the same context of *persevere* are closely connected with the Christian context or indicate difficulties and/or strong determination. Having access to the subtle usage differences in the semantic prosody and collocational behavior between the two synonyms, as illustrated in this article, will enable English learners to select an appropriate word for each particular context, thereby improving accuracy and proficiency in their use of L2 English.

The findings of this study can benefit English language teachers in a number of ways. First, in consultation with the present-study results, they will provide students with a clear explanation on the major distinction between *persist* and *persevere* with regard to the degree of formality, collocations, and semantic preference/prosody. Second, they can promote students' inductive learning of vocabulary skills by introducing the common collocates of each synonym through data-driven learning (DDL) (Friginal, 2018). For instance, in teaching the differences in synonym usage between *persist* and *persevere*, teachers can raise students' lexical awareness by preparing materials that include the selected common collocates that particularly co-occur with *persist* but not *persevere*, and vice versa. The materials should be based on authentic English, e.g., corpus-based information, on the grounds that students will be exposed to natural English in context, and then they will be asked to make some observations on the collocational tendencies of the target synonyms (Timmis, 2015). This way, students are expected to arrive at practical descriptions of language features, i.e., collocations, which are possibly more common than those found in general ELT materials (Gilquin, 2020).

Furthermore, teachers will also be able to examine the usage differences between other pairs of near-synonyms using a similar corpus-based approach, which will facilitate their teaching material development based on authentic language drawn from corpus data (Webb and Nation, 2017). In other words, English teachers can enhance their vocabulary instruction through authentic English as represented in corpora.

### **The Conflict of Interest Statement**

In line with the statement of Committee on Publication Ethics (COPE), I/we hereby declare that I/we had no conflicting interests regarding any parties of this study.

### **Acknowledgements**

This study was supported by Language Institute, Thammasat University Research Fund, Contract No. 2/2021.

### **References**

- Barnbrook, G, Mason, O. & Krishnamurthy, R. (2013). *Collocation. Applications and implications*. Palgrave Macmillan.
- Cheng, W. (2012). *Exploring corpus linguistics. Language in action*. Routledge.
- Chung, S-F. (2011). A corpus analysis of 'create' and 'produce'. *Chang Gung Journal of Humanities and Social Sciences, 4(2)*, 399-425.
- Conklin, K., & Schmitt, N. (2008.March). Formulaic sequences: Are they processed more quickly than non-formulaic language by native and non-native speakers? *Applied Linguistics, 29*, 72-89. <https://doi.org/10.1093/applin/amm022>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34(2)*, 213-238. <https://doi.org/10.2307/3587951>
- Crawford, W. J. & Csomay, E. (2016). *Doing corpus linguistics*. Routledge.
- Davies, M. (2020). *The corpus of contemporary American English*.
- Firth, J. R. (1957). *Papers in linguistics, 1934-1951*. Oxford University Press.
- Flowerdew, L. (2012). *Corpora and language education*. Palgrave Macmillan.
- Friginal, E. (2018). *Corpus linguistics for English teachers*. 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.
- Gilquin, G (2020). Using corpora to foster L2 construction learning: A data-driven learning experiment. *International Journal of Applied Linguistics. Special Issue*. <https://doi.org/10.1111/ijal.12317>
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Jackson, H. & Amvela, E. (2007). *Words, meaning, and vocabulary*. Cassel.
- 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.
- Kuiper, K., Columbus, G., & Schmitt, N. (2009). The acquisition of phrasal vocabulary. In S. Foster-Cohen (Ed.), *Language acquisition* (pp. 216-240). Palgrave Macmillan.
- Lewis, M. (1997). *Implementing the lexical approach*. Language Teaching Publications.
- Lewis, M. (2000). *Teaching collocation: Further development in the lexical approach*. Commercial Colour Press Plc.
- Lindquist, H. & Levin, M. (2018). *Corpus linguistics and the description of English*. Edinburgh University Press.
- Louw, B. (1993). Irony in the text or insincerity in the writer? In M. Baker et al. (Eds), *Text and technology* (pp. 157-176). John Benjamins.
- Longman dictionary of contemporary English (2014). Pearson Education.

- McCarthy, M. and O'Dell, F. (2005). *English collocations in use*. Cambridge University Press.
- Moon, R. (2010). What can a corpus tell us about lexis? In A. O'Keefe & M. McCarthy, (Eds.), *The Routledge handbook of corpus linguistics* (pp. 197-211). Routledge.
- Nelson, M. (2006). Semantic association in business English: A corpus-based analysis. *English for Specific Purposes*, 25, 217–34. <https://doi.org/10.1016/j.esp.2005.02.008>
- Partington, A. (1998). *Patterns and meanings. Using corpora of English language research and teaching*. John Benjamins.
- Partington, A. (2004). “Utterly content in each other's company”: Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1), 131-156.
- Petcharat, N. & Phoocharoensil, S. (2017). A corpus-based study of English synonyms: ‘appropriate’, ‘proper’, and ‘suitable’. *LEARN Journal: Language Education and Acquisition Research Network*, 10(2), 90-99.
- Phoocharoensil, S. (2010). A corpus-based study of English synonyms. *International Journal of Arts and Sciences*, 3(10), 227-245.
- Phoocharoensil, S. (2020). A genre and collocational analysis of ‘consequence’, ‘result’, and ‘outcome’. 3L: Language, Linguistics, Literature. *The Southeast Asian Journal of English Language Studies*, 26(3), 1-16. <http://dx.doi.org/10.17576/3L-2020-2603-01>
- Schmitt, N. (2010). *Research vocabulary. A vocabulary research manual*. Palgrave Macmillan.
- Selmistraitis, L. (2020). Semantic preference, prosody and distribution of synonymous adjectives in COCA. *GEMA Online® Journal of Language Studies*, 20(3), 1-18. <http://dx.doi.org/10.17576/gema-2020-2003-01>
- Sinclair, J. (1991). *Corpus, concordance, collocation: Describing English language*. Oxford University Press.
- Sinclair, J. (2004). *Trust the text*. London: Routledge.
- 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*. Routledge.
- Thornbury, S. (2002). *How to teach vocabulary*. Longman.
- Timmis, I. (2015). *Corpus linguistics for ELT*. Routledge.
- Webb, S. & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.
- Xiao, R. & McEnery, T. (2006). Collocation, semantic prosody and near synonymy: A cross linguistic perspective. *Applied Linguistics*, 27(1), 103-129. <https://doi.org/10.1093/applin/ami045>

---

### Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (CC BY-NC-ND) (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).