

Multiword Units and Synonymy: Interface between Collocations, Colligations, and Semantic Prosody

Supakorn Phoocharoensil
supakorn.p@litu.tu.ac.th

Language Institute of Thammasat University, Thailand

ABSTRACT

Appropriate use of near-synonyms in English often poses difficulty for learners. This study aims to compare and contrast the near-synonyms *predict* and *foresee*, focusing on genres, collocations, colligations, and semantic prosody which specifically characterize each synonymous verb. The study consulted the Corpus of Contemporary American English (COCA), i.e. the largest corpus representing American English (Davies, 2020), for the distribution of the target synonyms across genres and the top-20 noun collocates the MI-score of which is 3 or above. The total frequency of *predict* in the corpus is far higher than that of *foresee*. Of all the eight genres of COCA, *predict* is the most common in academic texts, whereas the highest number of occurrences of *foresee* is found in webpages. Interestingly, both target synonyms are not characteristic of colloquial English as their frequencies in speaking, fiction, and TV and movie subtitles are relatively low. The data from COCA also reveal that both synonymous verbs are common in written English, with *predict* being far higher in frequency. Although sharing certain object noun collocates, it is indicated through semantic prosody, i.e. occurrences of adjacent collocates and surrounding lexical items, that *foresee* is associated with negative connotations, while *predict* does not primarily express adversity. As near-synonyms of each other, the target words share colligational patterns. However, two corpus-informed syntactic structures, i.e. *foresee* + *somebody* + *V.ing* and *predict* + *somebody* + *to-infinitive*, clearly distinguish between both verb synonyms. It is concluded from the corpus data that *predict* and *foresee* are near-synonyms rather than absolute synonyms since they clearly differ in connotations and collocational/colligational patterns.

Keywords: synonymy; collocation; colligation; semantic prosody; semantic preference

INTRODUCTION

Vocabulary description and instruction have tended to concentrate on single words, often regarded as the basic lexical unit which is convenient to teach and incorporate into materials (Schmitt, 2010). However, it is becoming increasingly important that the vocabulary unit larger than an individual word requires far more attention in English Language Teaching (ELT). In actuality, vocabulary “behaves not as single words...but rather has a strong tendency to occur in multiple word phraseological units” (Schmitt, 2010, p. 117). This notion popularizes the significance of ‘formulaic language’ or ‘multiword units’, defined as “a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated; that is, stored and retrieved whole from memory at the time of use...” (Wray, 2002, p. 9). In a nutshell, a multiword unit, e.g. *a powerful argument* or *at the top of*, refers to “a vocabulary item which consists of a sequence of two or more words which semantically and/or syntactically forms a meaningful unit” (Moon, 1997, p. 43).

Research in vocabulary has indicated the widespread existence of formulaic language in spoken and written English, suggesting that formulaicity is a common linguistic phenomenon (Conklin and Schmitt, 2008). Different types of multiword units exist in English and they are referred to in a variety of terms, each of which has its own characteristic definition,

e.g. chunk (O’Keeffe, McCarthy, and Carter, 2007), lexical bundle (Biber, 2006), lexical phrase (Nattinger and DeCarrico, 1992), collocation (Lewis, 1997), and formulaic sequence (Wray, 2002), etc. No matter how it is classified, knowledge of formulaic language is an essential element of language users’ proficiency. Since it is an undeniable fact that words do not occur in isolation, their meanings can be understood via an exploration of “the different lexical relations they have with one another” (Szudarski, 2018, p. 43).

As the main purpose of this study is to investigate synonymy, a type of lexical relation, relevant kinds of multiword units were taken into the researcher’s consideration in analyzing the similarities and differences between the near-synonyms *predict* and *foresee*. The two synonymous verbs were selected for this study because *predict* occurs in the first 3,000 words of written English (Longman dictionary of contemporary English, 2014) and also listed in Coxhead (2000)’s Academic Word List or AWL (2000), and *foresee* is labeled C1 according to the CEFR, indicative of the advanced English level. More precisely, the present study focuses on the typical collocations, colligations, and semantic prosody/preference that are specifically associated with each synonym. Analyzing a large amount of corpus data in the process of near-synonym discrimination, the study aims to reveal principal differences in the phraseological patterns characterizing each target synonym on the grounds that no perfect synonyms, i.e. the near-synonyms that are interchangeable in all contexts, are existent in English (Carter, 2012; Wongkhan and Thienthong, 2020).

LITERATURE REVIEW

COLLOCATIONS

One of the most well-known types of multiword units worth introducing to English learners is ‘collocation’. Collocation is defined as “the phenomenon surrounding the fact that certain words are more likely to occur in combination with other words in certain contexts” (Baker, Hardie, and McEnery, 2006, p. 36). Lewis (1997) refers to collocation as “the readily observable phenomenon whereby certain words co-occur in natural text with greater than random frequency” (p. 8). Collocations are considered fixed and recurrent word-combinations (Cowie and Howarth, 1996). For instance, the noun *dog* and the verb *bark* are common collocates, i.e. words that frequently appear together, of each other due to their high probability of co-occurrence in texts, whereas the combination *cats bark*, which might make some sense, is a rarity in English (O’Keeffe, McCarthy, and Carter, 2007). Such a noun + verb collocation as *dogs bark* is sometimes referred to as a ‘lexical collocation’, which consists of only content or lexical words, e.g. nouns, verbs, adjectives, adverbs (Benson, Benson, and Ilson, 2009).

There are different classifications of collocations. One classification that is pedagogically useful is Hill’s (2000): strong, medium-strength, and weak collocations. An example of weak collocations is *good boys* since the adjective *good* can be combined with a number of words, as in *good houses*, *good books*, *good watches*, etc. On the other hand, the combination *dulcet tones* is a strong collocation in the sense that *dulcet* infrequently co-occurs with other nouns other than *tones*. Weak and strong collocations should not be the focus of ELT because weak collocations are easy and not problematic for learners’ comprehension and production, while strong collocations are low in frequency and it is unlikely that learners will encounter them in daily life. It is the medium-strength collocation that should receive ELT practitioners’ attention as this particular kind of collocation is high in frequency and its component co-selects a limited number of words with which it collocates. For instance, the adjective *severe* can collocate with a restricted number of nouns, such as *discipline*, *punishment*, *winter*, so their adjective+noun combination is regarded as a medium-strength

collocation. Learning such medium-strength collocations extends learners' collocational knowledge (Timmis, 2015).

Collocation is necessary for English learners to produce natural-sounding speech and writing. English, like other languages, contains a myriad of formulaic language, lexical chunks, and collocations. For a learner, choosing the right collocation "...will make his or her speech and writing sound much more natural, more native-speaker-like..." (Oxford collocations dictionary, 2009, p. v). Knowledge of collocations facilitates fluent communication in L2 English as it has been revealed that non-native speakers of English read and understand formulaic sequences significantly faster than non-formulaic chunks (Conklin and Schmitt, 2008). Furthermore, collocations are arbitrary, which is challenging for learners with limited L2 exposure or experience to master. For example, *heavy* and *strong*, both of which are adjectives describing a high degree or level, typically collocate with different nouns, as in *heavy rain* (rather than **heavy wind*) and *strong wind* (rather than **strong rain*). The significance of collocations in ELT encourages teachers to shift the emphasis from single word instruction to chunk integration in their vocabulary lessons. In summary, for L2 English learners, learning collocations "is not a luxury if anything above a survival level mastery of the language is desired, since collocation permeates even the most basic, frequent words" (O'Keefe, McCarthy, and Carter, 2007, p. 60).

It is important to note that collocations can be used to distinguish near-synonyms. While near-synonyms are similar in cognitive or denotational meanings, they are not always collocationally interchangeable (Xiao and McEnery, 2006), meaning that two synonyms tend to co-occur with different collocates. Learners who assume that two synonyms can be used with exactly the same sets of collocates are likely to produce L2 collocational patterns that sound unnatural since "[e]ven the slightest adjustments to the collocation – by substituting one of its components for a near-synonym...turns the text into non-standard English" (Thornbury, 2002, p. 7). A number of studies investigating subtle differences between English near-synonyms focus on collocation as a distinguishing criterion (e.g. Aroonmanakun, 2015; Chung, 2011; Crawford and Csomay, 2016; Jirananthiporn, 2018; Partington, 1998).

Most recent synonym studies analyzed data drawn from language corpora. Aroonmanakun (2015) pointed out the benefits of corpora in synonym differentiation, looking at the top 100 collocates of *quick* and *fast* with Mutual Information (MI) scores of at least 3 in the Corpus of Contemporary American English (COCA). Both synonyms seem to co-occur with different collocating nouns in that *quick* often refers to the action having been done or responded to in a short time, e.g. *quick answer*, *quick breakfast*, *quick reference*, *quick tip*, *quick visit*, while *fast* indicates the manner of movement rather than a short period of time, e.g. *fast acceleration*, *fast attack*, *fast boat*, *fast ride*, *fast tempo*. Crawford and Csomay (2016), in examining *equal* and *identical* in COCA, reported that corpus data allow for more in-depth observation of both near-synonyms. In terms of frequency, occurrences of *equal* (20,480 times) are higher than those of *identical* (8,080 times). With regard to noun collocates, *equal* has more tendency to be combined with abstract concepts, e.g. *opportunities*, *rights*, and *protection*, while *identical* typically collocates with concrete nouns, e.g. *twins*, *houses*, and *items*. In a similar vein, Jirananthiporn (2018) made a distinction between *problem* and *trouble*, relying on the data from COCA. It was found that *problem* is relatively more formal than *trouble* across five different text types. The difference between the two noun synonyms was made clearer when it comes to their collocations. More specifically, the verbs commonly collocating with *problem* are indicative of high formality and written discourse, e.g. *alleviate*, *eliminate*, *rectify*, *acknowledge*, *analyze*, *identify*, *address*, *confront*, *exacerbate*. In contrast, the verb collocates of *trouble* generally have a lower degree of formality, e.g. *ask*, *expect*, *like*, *mean*, *invite*, *want*, *spell*, *start*.

The next section will take into consideration the concept of ‘colligation’, which focuses on the co-occurrence of grammatical elements in the investigation of near-synonyms.

COLLIGATIONS

Colligation is a multiword unit that is closely related to collocation. While collocation deals with the lexical company a word or phrase is associated with, colligation refers to the co-selection of certain word classes or structural patterns surrounding the target or node words (Cheng, 2012). In other words, colligation is “the grammatical environment in which a word usually occurs. This can encompass parts of speech, tense, voice, or a word’s particular position in the sentence” (Flowerdew, 2012, p. 320). In an analysis of colligations, researchers explore patterns of co-occurring words vis-à-vis grammatical categories and syntactic relationships (Cheng, 2012). For instance, perception verbs, such as *hear*, *notice*, *see*, and *watch*, usually occur with either bare infinitives or *-ing* forms, as in *see + somebody + verb-infinitive/ verb-ing* (Nattinger and DeCarrico, 1992). Another example is the phrase *so far*, which is highly associated with the present perfect tense. More interestingly, data from British National Corpus (BYU-BNC) demonstrates that *so far* can occur in various positions, e.g. sentence-finally, sentence-initially, and within-verb-phrase. Learners can gain a clearer understanding of the discourse structure through “information about the different positions of colligations within specific parts of texts...” (Szudarski, 2018, p. 81).

It is worth noting that frequency and statistics are often used in an investigation of colligations. To illustrate, the phrase *I was wondering* can co-occur with the conjunctions *if* and *whether*, the wh-word *how*, and the preposition *about*. The search in the BYU-BNC corpus shows that *I was wondering if* occurs with the highest frequency, and *I was wondering whether* has the strongest collocational relationship as determined by the Mutual Information (MI) value despite its lower frequency than the former (Jones and Waller, 2015, p. 86). Another function or grammatical word that often constitutes colligations is the preposition. The colligation made up of a preposition is also known as a grammatical collocation (Benson, Benson, and Ilson, 2009). Like other grammatical elements, prepositions as a key component of colligation play a crucial role in enhancing native-like fluency. Incorrect choice of prepositions results in unnatural L2 English production. Native speakers of English will reject violations of collocability arising from misuse of prepositions, such as **accuse somebody on a crime*, as opposed to *accuse somebody of a crime* (Benson, Benson, and Ilson, 2009. P. XIX).

Colligations are the focus of several studies on synonymy (e.g. Ly and Jung, 2015; Phoocharoensil, 2010; Tognii-Bonelli, 1993). Synonymous words were compared and contrasted in detail using colligational patterns as a major distinguishing criterion. Tognini Bonelli (1993) reveals that approximately 99% of occurrences of *actual* follows the definite article *the*, whereas only 15% of the tokens of the near-synonym *real* co-occur with *the*. Phoocharoensil (2010) indicates that although the synonyms *ask*, *plead*, *request*, and *appeal*, partly share some grammatical patterns, there seem to be certain colligational patterns that are specifically characteristic of each individual synonym. For example, while the pattern *ask + somebody + to-infinitive* is very common, the remainder of the near-synonyms requires other patterns, e.g. *plead + with + somebody + to-infinitive*, *request + something + from + somebody*, *appeal + to + somebody + for + something*. Ly and Jung (2015) also found that even though the synonyms *cute*, *pretty*, and *beautiful* share some grammatical patterns, such as *be + cute/pretty/beautiful + to-infinitive*, *pretty* frequently appear in the syntactic pattern *not + pretty to watch* in describing unpleasantness and ugliness.

SEMANTIC PREFERENCE AND SEMANTIC PROSODY

In addition to the concepts of collocation and colligation, ‘semantic preference’ and ‘semantic prosody’ are multiword units that involve lexical patterning that demonstrates tendencies of words to form a chunk with one another and “become extended units of meaning” (Szudarski, 2018, p. 85). The semantic preference of a vocabulary item refers to “the tendency for lexical items to be restricted to identifiable semantic fields” (Cheng, 2012, p. 114). Semantic preference is defined as “the relation, not between individual words, but between a lemma or word-form and a set of semantically related words” (Stubbs, 2001, pp.65-66). Similarly, according to Flowerdew (2012), semantic preference, also known as semantic association, has a close connection with collocation as can be seen in the verb *cause*, which has a preference for noun collocates denoting diseases, e.g. *cancer*, *heart disease*.

In examining semantic preference, lexical items can be grouped according to semantic similarity (McEnery and Hardie, 2012). Sinclair (1999) exemplifies the semantic preference of *the naked eye* in relation to visibility, including words such as *seen*, *visible*, and *perceived*. Baker, Hardie and McEnery (2006) also clarify how words are classified on the basis of meaning similarity looking at the word *rising* in the British National Corpus (BNC), which is inclined to be combined with noun collocates related to work and money, such as *incomes*, *prices*, *wages*, *earnings*, *unemployment* (p. 144). Semantic preference, furthermore, also occurs with phrases, as in the phrase *glass of*, which typically collocates with a set of lexical items denoting drinks, such as *sherry*, *water*, *milk*, *lemonade*, *champagne*, etc. (p. 144).

Certain principal characteristics of semantic preference should be noted. First, the grammatical environment in which lexical items occur determines semantic preference for a particular domain. By way of illustration, the verb *cause* tends to co-occur with different lexical sets depending on the type of object with which it precedes. To be more precise, when *cause* is followed by a single object, the object noun is often concerned with illness, e.g. *cancer*, *heart disease*. However, where *cause* has double objects, the second object frequently conveys unpleasant or undesirable feeling, e.g. *convenience*, as in *causes them inconvenience* (Partington, 2004). The second characteristic concerns the varying degree of semantic preference based on the type of corpora being consulted. To put it another way, the corpus type on which an analysis is based determines the semantic preference of a lexical item. To support this notion, Nelson (2006) shows the semantic preference of the adjective *global* which differs from one corpus to another. For example, in a one-million-word business English corpus (BEC), *global* is rich in semantic sets such as *global products* and *global economic indicators*. On the other hand, *global* is limited to two semantic sets related to climate, as in *global warming*, and people, as in *global viewer*, as noticed in a general corpus like the British National Corpus (BNC).

In addition to semantic preference, corpus linguists are often interested in ‘semantic prosody’ in their study of phraseology, i.e. a concept closely connected with semantic preference and collocation. A classic definition of semantic prosody is “a consistent aura of meaning with which a form is imbued by its collocates” as provided by Louw (1993, p. 157). Semantic prosody relates to “attitudinal or evaluative meanings which result from a word’s co-occurrence with specific collocations” (Szudarski, 2018, p. 86). It is obvious that this concept, dealing with recurrent phraseological patterns, is primarily concerned with connotation. That is, lexical words can have a negative or positive semantic prosody “if they typically co-occur with units that have a negative or positive meaning” (McEnery and Hardie, 2012, p. 136). To illustrate, the verb *happen* has a tendency to co-occur with subject nouns that can be evaluated as negative, e.g. *accident*, *disaster*, etc., despite the fact that this verb can sometimes appear with subject nouns whose meanings are positive, e.g. *miracle* (Timmis, 2015). Semantic prosody is also noticeable in phrasal verbs like *break out*, meaning ‘occur’; the corpus data

indicate that there is a high likelihood of *break out* co-occurring with subject nouns denoting physical violence, e.g. *blaze, fight, fire, hostilities, violence, war* (Ruehleemann, 2010).

Hunston (2002) summarizes the main features of semantic prosody as follows. First, semantic prosody expresses the connotation of a word in association with its neighboring words. A positive or negative connotation is not easily spotted from a word being in isolation but from its combination with others. One of the pieces of concrete evidence is *bordered on* and *bordering on*, which typically collocate with words indicating something undesirable, such as *antagonism, alcoholism, blackmail, carelessness, chaos, contempt, conspiracy, cruelty, cynicism*, etc. (Sinclair, 2004). Second, the connotation represented by the semantic prosody is not always accessible to native speakers' intuition. Even though semantic prosodies may occasionally confirm native speakers' intuition regarding certain connotations, the positive or negative associations of a word or phrase could be inaccessible from native speakers' conscious knowledge. Hunston (2002) gives *set in* as an example of lexical items whose connotative meaning 'something bad starts to happen' is normally beyond English native speakers' intuition. In other words, native speakers will become enlightened with respect to some lexical usage when they are presented with a large number of examples or words or phrases obtained from language corpora. This is considered to be the third feature of semantic prosody in that considerable linguistic data from concordance lines representing authentic English allows for better understanding of connotation or more convincing generalizations of some particular lexical items.

With regard to the focus of this study, i.e. near-synonym discrimination, it is advisable that semantic preference, semantic prosody, and collocations be taken into account as they are interdependent in the study of subtly different meanings between near-synonyms (Flowerdew, 2012). A clear example of research study on synonyms is Partington (1998), searching for similarities and differences in meaning usages among the adjective synonyms *sheer, pure, complete*, and *absolute*. Through an investigation of the semantic properties of the common noun collocates of *sheer* in academic and newspaper corpora, it was discovered that *sheer* co-occurs with the following semantic categories of noun collocates:

- (a.) collocates that express magnitude, weight, or volume, e.g. *the sheer volume of reliable information, the sheer weight of numbers, the sheer number of artefacts, the sheer scale of the selling*
- (b.) collocates that express force, strength, or energy, e.g. *the sheer physical energy, the sheer force of his personality, the sheer pressure on existing farmlands*
- (c.) collocates that express persistence, e.g. *sheer persistence, the sheer irreversibility, the sheer hard work, the sheer resolve on the right side*
- (d.) collocates that express strong emotion, e.g. *the sheer joy of playing the game, sheer jealousy, its sheer inspiration, sheer terror*

In comparison with *sheer*, the other target synonymous adjectives, i.e. *pure, complete*, and *absolute*, have their own specific patterns of occurrence. Apparently although these synonyms are similar in core meanings, they can never perfectly replace *sheer* in all contexts of use due to the fact that "every lexical item in the language has its own individual and unique pattern of behavior" (Partington, 1998, p. 46). For instance, *pure* can be used to describe 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*, while *sheer* hardly expresses such particular senses. It should also be noted that *complete* is not a 'strict' or perfect synonym of *sheer* either as the following collocates of *complete*, e.g. *abandonment, absence, destruction, transfer, withdrawal*, do not express multitude or force. Just like *pure* and *complete*, the synonym *absolute* has a sense of 'unchallenged' or 'unquestioned', as in *absolute control of*

their schools or *absolute control of information*, a sense that is absent in *sheer*. To sum up, collocations and semantic preference/prosody that each synonym is associated with can identify the differences between pairs or sets of near-synonyms.

In the present study, the near-synonyms *foresee* and *predict* were explored with an emphasis on the genres in which they occur, colligations, and their collocational patterns that account for the semantic prosody/preference characterizing each synonymous verb. It is hypothesized that the two synonyms differ in the object noun collocates that they are often combined with. The difference in collocations between synonyms, in turn means the difference in connotations reflected through the semantic prosody.

METHOD

DATA COLLECTION PROCEDURE

The research procedure involves the following steps. First of all, the Longman Dictionary of Contemporary English, 6th Edition (LDOCE), which is a well-known learner dictionary offering thousands of corpus-based examples, grammar and error notes to facilitate English learners' understanding of English words, was consulted for the definitions, sample sentences, and usage notes of the synonyms *foresee* and *predict*. The information from the learner dictionary was then compared with that derived from the Corpus of Contemporary American English (COCA). COCA was selected because of its huge size and various genres, representative of general English, made up of over one billion words, with around 20 million words being included on an annual basis from 1990-2019. Moreover, COCA has been a reliable, well-known source of authentic English, most widely used among corpus linguists as well as ELT researchers and teachers (Friginal, 2018; Phoocharoensil, 2020). The newest version of COCA, launched in March 2020, consists of texts of eight different genres, including five fundamental genres, i.e. spoken, fiction, popular magazine, newspaper, and academic texts, and three new genres, i.e. TV and movie subtitles, blogs, and webpages (Davies, 2020).

Second, common noun collocates of *foresee* and *predict* were extracted from COCA on the basis of the strength of collocational association measured by the Mutual Information (MI) score in conjunction with the frequency. High MI scores mean a strong association of collocation. Nevertheless, it is noteworthy that collocation extraction based on MI score alone may cause a problem as some collocations with very high MI scores are often found to be low in entire frequency in a corpus. In other words, reliance on MI scores in collocation study can result in a list of rare occurrences due to the fact that the MI value "promotes rare words" in ranking the collocates (Barnbrook, Mason, and Krishnamurthy, 2013, p. 67). Such odd or rare collocations, therefore, are not worth introducing to students since rarely will they meet those collocations again in everyday life (Schmitt, 2010). For the present study, the collocational strength was rated based on both the frequency of noun collocates and the MI scores in order to ensure that recurrent, frequent collocates would be extracted. Consequently, the top-20 high-frequency noun collocates presented in COCA whose MI score is ≥ 3 , i.e. the significance value for collocational strength (Cheng, 2012), were selected.

In the next step, the noun collocates of *foresee* and *predict* were grouped according to the semantic preference; that is, collocates that share similarities in meaning were assigned to the same set or category. The common colligational patterns and semantic prosody, viewed as a subset of semantic preference (Flowerdew, 2012), of the two verb synonyms, was also examined from 100 concordance lines of each synonym randomly selected from COCA.

RESULTS AND DISCUSSION

The definitions of *foresee* and *predict* given in LDOCE (2014) are as follows:

foresee [transitive] to think or know that something is going to happen in the future

e.g. I've put your name on the list and I don't ***foresee*** any problems.
The disaster could not have been ***foreseen***.

foresee that

e.g. Few analysts ***foresaw*** that oil prices would rise so steeply.

foresee what/how etc

e.g. No one foresaw what he was planning.

(LDOCE, 2014, p. 717)

predict [transitive] to say that something will happen, before it happens

e.g. Sales were five percent lower than ***predicted***.

predict that

e.g. Newspapers ***predicted that*** Davis would be re-elected.

predict whether/what/how etc

e.g. It is difficult to ***predict what*** the long-term effects of the accident will be.

be predicted to do something

e.g. Unemployment ***is predicted to increase*** to 700,000 by the end of the year.

(LDOCE, 2014, p. 1421)

According to the definitions cited above, *foresee* and *predict* seem to be synonymous since they both generally share the denotative meaning of 'say or know that something will happen in the future'. Furthermore, two colligational patterns overlap between the two synonymous transitive verbs, i.e. *forecast/predict that* and *forecast/predict + wh-word*, except that *predict* can also precede the conjunction *whether*. However, the phraseological pattern *be predicted to do something* appears to be associated with *predict* as opposed to *foresee*. Interestingly, unlike some other common words in LDOCE, no information on typical lexical collocations is available for either word. The dictionary does not provide any difference in connotations each synonym conveys. This was why the current study, in an attempt to explore corpus data from COCA, aimed to uncover subtle differences in terms of semantic prosody/preference informed by the lexical items surrounding the target synonyms.

Next, the findings with regard to the distribution of *foresee* and *predict* across genres in COCA will be shown and discussed.

TABLE 1. Distribution of the Synonyms *Foresee* and *Predict* across Genres according to Frequency

Genre	<i>Foresee</i>		Genre	<i>Predict</i>	
	Frequency	Per million		Frequency	Per million
webpages	325	2.62	academic texts	5,132	42.84
blogs	326	2.53	magazines	3,094	24.54
academic texts	288	2.40	webpages	2,673	21.51
magazines	301	2.39	blogs	2,742	21.32
newspapers	290	2.38	newspapers	2,114	17.36
spoken	297	2.35	spoken	2,156	17.09
fiction	161	1.36	fiction	644	5.44
TV and movie subtitles	105	0.82	TV and movie subtitles	664	5.18
Total	1,767		Total	16,511	

In Table 1, it is clearly seen that *predict* (16,511 tokens) occurs with much higher frequency than *foresee* (1,767 tokens) in all the genres in COCA. In particular, *predict* is the most frequent in academic texts (42.84 per million), which implies its formality, followed by magazines (24.54 per million). It is worth noting that the frequencies in webpages (21.51 per million) and blogs (21.32 per million) are close. This similarity in frequency is also found in newspapers (17.36 per million) and spoken language (17.09 per million). Its frequency is lowest in TV (5.44 per million) and movie subtitles (5.18 per million). In contrast, *foresee*, whose occurrences are substantially lower than those of *predict*, has its highest frequency in two related genres, i.e. webpages (2.62 per million) and blogs (2.53 per million) respectively. Its per-million-word frequencies in academic texts (2.40 per million) and magazines (2.39 per million) are almost equal. Likewise, there are similarities in frequency between *foresee* in newspapers (2.38 per million) and that in spoken language (2.35 per million). The lowest frequency of *foresee* is in TV and movie subtitles, which is consistent with the finding of *predict*. Both synonyms seem to be associated with formality due to their relatively high frequency in written English, i.e. academic texts, webpages, and blogs, and their infrequent occurrence in spoken language, e.g. fiction and TV/movie language.

TABLE 2. Noun Collocates of *foresee* and *predict* from COCA

Rank	<i>Foresee</i>			<i>Predict</i>		
	Noun collocate	Frequency	MI-Scores	Noun collocate	Frequency	MI-Scores
1	future	136	4.51	future	1057	3.91
2	consequence	63	4.87	outcome	871	5.08
3	possibility	58	4.40	behavior	634	3.41
4	outcome	33	3.82	performance	518	3.25
5	difficulty	30	4.15	weather	425	4.07
6	danger	28	3.72	score	301	3.08
7	scenario	27	4.41	achievement	235	3.91
8	potential	19	3.60	earthquake	141	4.31
9	circumstance	17	3.16	accuracy	131	4.20
10	happening	17	5.35	satisfaction	128	3.75
11	rise	15	3.20	certainty	111	4.73
12	collapse	15	4.26	likelihood	98	4.04
13	disaster	14	3.17	demise	86	4.93
14	revolution	13	3.05	variance	84	4.15
15	decline	12	3.61	probability	79	3.52
16	expansion	10	3.30	earnings	76	3.19
17	inability	10	4.52	comprehension	76	4.25

18	harm	9	3.31	collapse	70	3.02
19	recession	7	3.06	mortality	68	3.49
20	shortage	7	3.07	turnout	55	4.14

Based on the total frequency and MI scores, the top-20 high-frequency noun collocates of *foresee* and *predict* were extracted. As shown in Table 2, while the two synonyms seem to share three object nouns, i.e. *future*, *outcome*, and *collapse*, the lists above need to be interpreted with caution due to the fact that there are actually many other nouns that can collocate with both *foresee* and *predict*, e.g. *problem*, *consequence*, *disaster*, *trend*, etc., but were not included in the lists because their frequencies are lower than the ones being extracted or their MI score does not reach 3, i.e. the significant level of collocational strength (Cheng, 2012).

The semantic preference of the target synonyms was then examined to categorize the noun collocates on the basis of their semantic properties. In corpus-based studies of synonymy, semantic preference is normally determined by the semantic relations between lexical items and their neighboring words. In other words, a list of common collocates helps identify the range of word associations and the semantic relationships among collocates, as seen in Table 3 (Szudarski, 2018).

TABLE 3. Semantic Preference of Noun Collocates of *Foresee*

1. RESULT	consequence, outcome
2. POSSIBILITY/FUTURE SITUATION	circumstance, future, happening, possibility, potential, scenario
3. SOMETHING ADVERSE OR UNPLEASANT	collapse, danger, decline, difficulty, disaster, harm, inability, recession, revolution, shortage
4. SOMETHING POSITIVE	expansion, rise

The nouns that commonly collocate with *foresee* were grouped into four main semantic sets based on their meaning similarities, as presented in Table 3. It is noteworthy that the majority of noun collocates convey negative or adversative meanings, i.e. *collapse*, *danger*, *decline*, *difficulty*, *disaster*, *harm*, *inability*, *recession*, *revolution*, and *shortage*. Only two nouns express positive meanings, i.e. *expansion* and *rise*. In addition, some other nouns denote events or situations in the future, e.g. *future*, *circumstance*, *happening*, and *scenario*, while some express possibility that something will occur or develop in a particular way, i.e. *possibility* and *potential*. Two nouns describe a future result, i.e. *consequence* and *outcome*.

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

1. POSSIBILITY	certainty, future, likelihood, probability, variance
2. MEASURABLE OUTPUT	accuracy, behavior, comprehension, earnings, outcome, performance, score, turnout
3. WEATHER	weather
4. SOMETHING ADVERSE OR UNPLEASANT	collapse, demise, earthquake, mortality
5. SOMETHING POSITIVE	achievement, satisfaction

The analysis of the semantic preference of the nouns collocating with *predict* shows that this verb shares some semantic properties with *foresee*, e.g. possibility, negative meanings, and positive meanings. The overlap in meaning and noun collocates between the two verbs is indicative of their status as near-synonyms. The largest semantic set of *predict* encompasses nouns denoting measureable output, i.e. *accuracy, behavior, comprehension, earnings, outcome, performance, score, and turnout*. The next sets include noun collocates that express possibility, i.e. *certainty, future, likelihood, probability, and variance*, and refer to weather, i.e. *weather*. A major difference that seems to distinguish *predict* from *foresee* lies in the nouns conveying negative or adversative meanings. That is, while a negative connotation is clearly associated with *foresee*, as can be seen from the common collocates of this verb in Table 3, adversity is not apparently characteristic of *predict*, which is more semantically neutral.

In order to gain a more profound understanding of the connotation related to the verb synonyms *foresee* and *predict*, 100 concordance lines of each verb were randomized from COCA and the semantic prosody was thoroughly analyzed. This closer investigation into the semantic prosody gives prominence to not only the collocates appearing right next to the target words but also other lexical items located away from the synonyms but within the same context. It was discovered that, as the noun collocates of *foresee* often signify negative connotations as previously stated, the existence of other words or phrases surroundings *foresee* is also a clear indication that this verb is inherently negative, whereas *predict* is not. Below are some examples of extracted concordance lines of *foresee* from COCA.

1. One may **foresee** that conception will be unlikely, or impossible...
2. I can almost **foresee** a new terrorist organization stemming from science.
3. I **foresee** a future in which phone-based e-commerce makes the leap from virtual to physical goods.
4. Do you **foresee** any awkwardness or weird questions coming up at the wedding itself?
5. I **foresee** a time when this will blow up on the puppet masters.
6. If things don't begin to change soon I **foresee** a time of violent social upheavals.
7. ... of my apprehensions and nervousness about what Nietzsche happened to **foresee**.
8. An effective property lawyer will be able to **foresee** potential problems that can rise involving the buyer and the seller.
9. ...foot for three whole days because she didn't originally **foresee** that foot injury would occur?
10. ...less to blame than anyone else for failing to **foresee** the recession or in responding too slowly to the glacial economic recovery.
11. They could **foresee** crimes before they even happened, allowing law enforcement to prevent them from ever...
12. I agree with this and **foresee** increasing investment by marketing divisions on "channel" discovery through the use biomarkers...
13. The architects did not **foresee** the enormous advances in electronic and pharmaceutical medicine, both of which themselves increase...
14. If President Obama is murdered, I **foresee** serious trouble in the streets.
15. The only drawbacks I **foresee** is the lack of variety of food and the greater travel distances required.

Through a close examination of the randomly selected concordance lines, up to 76 out of 100 lines of *foresee* hold a clear association with negative meanings. In particular, based on the 15 lines presented above, many object nouns are semantically adversative, e.g. *a new terrorist organization, any awkwardness or weird questions, awkwardness or weird questions, a time of violent social upheavals, my apprehensions and nervousness, potential problems,*

crimes, serious trouble in the streets, drawbacks, and the recession. Furthermore, when looking at the lexical items that are located in the same sentence in which *foresee* appears, unfavorable or unpleasant connotations are obvious, as demonstrated in longer units like *that conception will be unlikely or impossible, a time when this will blow up on the puppet masters, foot injury would occur, the lack of variety of food, the greater travel distances, and in responding too slowly to the glacial economic recovery.*

The negative sense that *foresee* imparts becomes noticeable when collocates and other surrounding words in the context are taken in account, which confirms the observation of Hunston (2002) that studying words in isolation alone cannot completely decode its subtle meaning; connotations or attitudinal meanings can be reflected by a corpus analysis of a word in combination with specific collocations (Szudarski, 2018). The discovery of negative connotations attached mainly to *foresee* seems to confirm a great benefit of corpus data that surpasses the information from learner dictionaries (Phoocharoensil, 2010) in that the latter source does not appear to pinpoint the adversity associated with this target verb. Put simply, the findings are consistent with McEnery and Hardie (2012) that the semantic prosody links the verb *foresee* to speakers' attitude "which may not be a single word, but may be given in the wider context" (p. 138).

It is of paramount importance to note that *foresee* can also co-occur with lexical items expressing positive meanings, e.g. *the enormous advances in electronic and pharmaceutical medicine* and *increasing investment*, as shown in the extracted concordance lines above, although this particular trend is outweighed by the negative one. Moreover, the context in which *foresee* occurs can be neutral in meaning as well, e.g. *a future in which phone-based e-commerce makes the leap from virtual to physical goods.* This corresponds to Timmis (2015), who claims that words do not always fall into a solely positive or negative semantic property. In fact, semantic prosody is viewed as "more of a probabilistic phenomenon rather than a plus-or-minus one" (p. 21). This means despite the fact that a word may be primarily associated with negative polarity, the semantic prosody of the word can sometimes be positive or neutral.

1. A lot of analysts **predict** prices could fall even further in the short term.
2. ...Academy has an interesting idea of trying to find out which sorts of test results **predict** a student's ability to answer other questions.
3. I **predict** their shareholders will be very happy indeed.
4. ...it does not **predict** which mental illness you're more likely to experience.
5. ...a circuit simulator with the model can quickly **predict** the performance of a circuit with multiple devices connected to it.
6. When trying to explain and **predict** a person's behavior, we typically refer to concepts such as the person's...
7. ...choose to end their lives, even though most doctors admit they cannot accurately **predict** life expectancy.
8. Would you bet your life on the ability of a computer model to accurately **predict** the future?
9. It is also hard to **predict** short-term consequences, although it is, perhaps paradoxically, much easier to predict the...
10. ...but this is like trying to **predict** the distance each wave will wash up the beach.
11. ...if you can't **predict** the outcome, well, that's a reflection on your capacity.
12. The second one is that when you **predict** the probabilities, it is much easier than predicting the total contribution of the impact.
13. Past risks do not **predict** future risks.
14. What they can do is **predict** temperature changes as a function of greenhouse gas emissions,...

15. Someday, we will be able to **predict** both the climate and the solar cycles, but not if the IPCC manages to...

As for *predict*, careful scrutiny of its semantic prosody indicates that unlike *foresee*, which often deals with negative meanings, the synonym *predict* tends to co-occur with a variety of noun collocates, which are neutral, positive, or negative in meaning. Put differently, *predict* is not particularly attached to adversity. As shown in the above sample concordance lines, *predict* occurs with only a few instances of negative words, e.g. *mental illness*, *future risks*, and probably *short-term consequences*, while only one token of positive semantic prosody is found in *I predict their shareholders will be very happy indeed*. In general, *predict* is usually combined with object nouns whose meanings are neutral, e.g. *prices*, *a student's ability*, *the performance of a circuit*, *a person's behavior*, *life expectancy*, *the future*, *the outcome*, *the probabilities*, *the distance*, *temperature changes*, or *the climate and the solar cycles*.

Now it is possible to make a clearer distinction between the two near-synonyms *foresee* and *predict* in terms of collocations and connotations as reflected by the semantic prosody. Although both verbs are close in cognitive meaning and share some object noun collocates, *foresee* is essentially connected with unpleasantness and adversity, whereas *predict* does not have such a negative implication. The findings regarding the adversative meaning related to the verb *foresee*, as indicated by the semantic prosody analysis, accords with past studies (e.g. Hunston, 2002; Partington, 2004; Ruehleemann, 2010; Sinclair, 2004). Such a difference in collocations and prosodic behavior between both synonyms confirms the fact that no absolute or perfect synonyms, i.e. those that are substitutable in all contexts of use, exist in language (Carter, 2012).

TABLE 5. Colligational Patterns of *Foresee* in COCA

Colligational patterns	Frequency	Examples from COCA	
foresee + noun	72	The administration's latest projections foresee \$6.4 trillion worth of deficits between 2013 and 2022.	
foresee + somebody/something + V.ing	15	I don't foresee her losing this case, but I don't see her winning it Do you foresee any awkwardness or weird questions coming up at the wedding itself?	
foresee + that-clause	9		
	that	7	It is able to foresee that serious sin seriously compromises the welfare of the sinner.
	(that)	2	...clearly there can be, and attempts to foresee those (such as Mr. Carr's book) are valuable and necessary.
foresee + wh-word	4		
	what	2	None of us can foresee what lies ahead for our economy.
	when	1	To me, that's something that I did not foresee when I first started on this venture many moons ago.
	how	1	I can't foresee how my values will change once there is a baby in the picture.
TOTAL	100		

According to Table 5, the colligational pattern *foresee* + *noun* appears to be the most frequent (71 tokens), based on the 100 concordance lines drawn at random from COCA. The second most frequent is *foresee* + *somebody/something* + *V.ing* (15 tokens), followed by *foresee* + *that-clause* (9 tokens), with the structure explicitly having *that* (7 tokens) higher in

frequency than the construction without *that* (2 tokens). The least frequent pattern is *foresee* + *wh-word* (4 tokens). The COCA data reveal the common grammatical pattern *foresee* + *somebody/something* + *V.ing*, which is not available in learner dictionaries, e.g. LDOCE. This colligational structure is analogous to the widely-known pattern *see* + *somebody/something* + *V.ing*, probably because both *foresee* and *see* share the sense of ‘vision’ or ‘sight’. This finding lends support to Phoocharoensil (2010) in that the information from learner dictionaries sometimes fails to cover all the common grammatical patterns that exist in native speakers’ language. Relying on data from language corpora representing authentic English is useful to English description as well as instruction as this can provides students and teachers with some further colligations that they do not see in dictionaries (O’Keeffe, McCarthy, and Carter, 2007).

TABLE 6. Colligational Patterns of *Predict* in COCA

Colligational patterns		Frequency	Examples from COCA
predict + noun		52	Nobody can predict market fluctuations.
predict + wh-word		24	
	what	12	Even the revolutionaries can't predict what will happen.
	how	7	Tests of enzyme levels may be able to predict how an individual patient will respond.
	which	2	I keep thinking that you can almost predict which fields will advance quickly and which will not.
	where	1	And so until there's clarification, it's hard to predict where that's going to go.
	who	1	It's really hard to predict who I end up keeping and who I lose.
	whether	1	Perhaps most surprising, strong math and reading skills were not found to predict whether kids would go on to complete college.
predict + that-clause		23	
	that	9	I predict that you will never see this charge backed by anything remotely resembling evidence.
	(that)	15	A lot of analysts predict prices could fall even further in the short term.
predict + somebody + to-infinitive		1	I don't predict them to win, or try not to,...
TOTAL		100	

The COCA data have shown four common colligational patterns of *predict*, as demonstrated in Table 6. Out of the 100 randomly-selected concordance lines, the most frequent pattern is *predict* + *noun* (52 tokens), followed by *predict* + *wh-word* (24 tokens) and *predict* + *that-clause* (23 tokens) respectively. With regard to the construction *predict* + *wh-word*, *what* is apparently the most common clause marker, followed by *how*. As for *predict* + *that-clause*, the structure in which the complementizer *that* is omitted occurs with higher frequency (15 tokens) than that containing *that* (9 tokens). However, it is worth noting that while LDOCE provides the pattern *be predicted to do something*, the corresponding active construction *predict* + *somebody* + *to-infinitive*, rather than the passive one, was found in COCA. The non-existence of the pattern *be predicted to do something* is probably because of the small size of data randomized from COCA. It is anticipated that more tokens of *predict* will yield additional findings, including the existence of this particular passive structure, as specified in the learner dictionary.

It is generalizable to a certain extent from the investigation into the colligational patterns of *foresee* and *predict* that the grammatical structures in which both verb synonyms are frequently used are not exactly the same. Even though they share three syntactic constructions, namely *foresee/predict* + *noun*, *foresee/predict* + *that-clause*, and *foresee/predict* + *wh-word*, each verb seems to have its own specific pattern not being shared by the other near-synonym, i.e. *foresee* + *somebody/something* + *V.ing* and *predict* + *somebody* + *to-infinitive*. Furthermore, it is also important to note that if we look in detail at a shared pattern like *foresee/predict* + *wh-word*, the colligational behavior of each synonym differs. It may be inferred from the selected concordance data that *whether* has more tendency to co-occur with *predict* rather than *foresee*, which is consistent with the LDOCE information. Such a difference in colligations between both near-synonyms confirms the findings of the previous studies that no synonyms are identical in grammatical patterns, and this syntactic usage helps differentiate synonyms, thus leading learners to proper use in the right grammatical context (e.g. Ly and Jung, 2015; Phoocharoensil, 2010; Tognii-Bonelli, 1993).

CONCLUSION

The present study focuses on distinguishing between the two verb synonyms *foresee* and *predict* in terms of genres, collocations, semantic prosody, and colligations. Regarding text types, both *foresee* and *predict* are characteristic of written language as *foresee* occurs with the highest frequency in webpages and *predict* is the most frequent in academic texts. Overall, the total frequency of *predict* is approximately 9.3 times higher than that of *foresee*. In particular, the occurrences of *predict* outnumber those of *foresee* in every genre. When collocations are taken into consideration, the two synonyms share a number of noun collocates, as indicated by COCA; however, only a few shared collocates are listed in this study on the basis of frequency and MI-scores, which explains why many noun collocates with which both verbs can co-occur are not included in the top-20 list. With noun collocates being grouped according to semantic preference, some semantic sets of the two synonymous verbs seem to overlap, i.e. ‘possibility’, ‘adversity’, and ‘positive sense’. Nonetheless, based on the semantic prosody analysis of COCA data, it is crucially important to see *foresee* frequently being combined with negative connotations, while *predict* appears neutral, not having such a negative implication. In addition, although *foresee* and *predict* have several grammatical patterns in common, it was revealed by the corpus-based data that some syntactic constructions are associated with an individual verb; that is, *predict* can occur in *predict* + *somebody* + *V.infinitive*, whereas *foresee* can be used in *foresee* + *somebody/something* + *V.ing*. More importantly, these two structures do not exist in learner dictionaries, e.g. LDOCE. The difference in colligational patterns, together with differences in other aspects, i.e. collocations, semantic prosody, and genres, contributes to a clearer, systematic discrimination between both near-synonyms.

RECOMMENDATIONS FOR FURTHER STUDIES

Future researchers who wish to conduct research on near-synonymy can also place an emphasis on the following aspects. First, in examining collocations and semantic prosody/preference, it is interesting to apply other statistical tests, apart from MI-score, to find collocational association without extracting potentially rare collocates, such as z-score, t-score, chi-squared, or log-likelihood tests (Gablasova, Brezina, and McEnery, 2017; Schmitt, 2010). This way, a different list of collocates may be obtained and a clearer picture of collocational behavior of both synonyms could be provided. Second, it is suggested that other pairs or groups of near-synonyms be investigated. Third, further studies may also look at the way English learners use a pair of synonyms in terms of collocations and colligations, in comparison with L1-speakers’

synonym production, concentrating on the variations between the two groups' synonym use. It is likely that the collocational and colligational behavior of synonyms that L2 learners employ will differ from what native speakers usually produce.

REFERENCES

- Aroonmanakun, V. (2015). 'Quick' or 'Fast': A Corpus-based Study of English Synonyms. *LEARN Journal: Language Education and Acquisition Research Network*. 8(1), 53-62.
- Baker, P., Hardie, A. & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Barnbrook, G, Mason, O. & Krishnamurthy, R. (2013). *Collocation. Applications and Implications*. London: Palgrave Macmillan.
- Benson, M., Benson, E. & Ilson, R. (2009). *The BBI Combinatory Dictionary of English. Your Guide to Collocations and Grammar*. Amsterdam/Philadelphia: John Benjamins.
- Biber, D. (2006). *University Language: A Corpus Study of Spoken and Written Registers*. Amsterdam and Philadelphia: John Benjamins.
- Carter, R. (2012). *Vocabulary. Applied Linguistic Perspectives*. London: Routledge.
- Cheng, W. (2012). *Exploring Corpus Linguistics. Language in Action*. London: 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). Formulaic Sequences: Are They Processed More Quickly Than Non-formulaic Language by Native and Non-native Speakers? *Applied Linguistics*. March. 29, 72-89.
- Cowie, A. & Howarth, P. (1996). Phraseological Competence and Written Proficiency. In G. Blue & R. Mitchell (Eds.), *Language and Education: British Studies in Applied Linguistics 11* (pp. 80-93). Clevedon: Multilingual Matters.
- 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.
- Davies, M. (2020). *The Corpus of Contemporary American English*.
- Flowerdew, L. (2012). *Corpora and Language Education*. London: Palgrave Macmillan.
- Friginal, E. (2018). *Corpus Linguistics for English Teachers*. London: Routledge.
- Gablasova, D, Brezina, V. & McEnery, T. (2017). Collocations in Corpus-based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence. *Language Learning*. 67(1), 155-179.
- 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.
- 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.
- Jones, C. & Waller, D. (2015). *Corpus linguistics for grammar*. London: Routledge.
- Kuiper, K., Columbus, G., & Schmitt, N. (2009). The Acquisition of Phrasal Vocabulary. In S. Foster-Cohen (Ed.), *Language Acquisition* (pp. 216-240). Hamshire: Palgrave Macmillan.
- Lewis, M. (1997). *Implementing the Lexical Approach*. Hove, English: Language Teaching Publications.
- Louw, B. (1993). Irony in the Text or Insincerity in the Writer? In M. Baker & E. Tognini-Bonelli (Eds), *Text and Technology* (pp. 157-176). Amsterdam: John Benjamins.
- Longman Dictionary of Contemporary English* (2014). Essex: Pearson Education.

- Ly, T. H. & Jung, C. K. (2015). A Corpus Investigation: The Similarities and Differences of ‘Cute’, ‘Pretty’ and ‘Beautiful’. *3L: The Southeast Asian Journal of English Language Studies*. 21(3), 125-140.
- McCarthy, M. & O’Dell, F. (2005). *English Collocations in Use*. Cambridge: Cambridge University Press.
- McEnery, Y. & Hardie, A. (2012). *Corpus Linguistics*. Cambridge: Cambridge University Press.
- Moon, R. (1997). Vocabulary Connections: Multi-word Items in English. In N. Schmitt & M. McCarthy, (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 40-63). Cambridge: Cambridge University Press.
- Nattinger, J. R. & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Nelson, M. (2006). Semantic Association in Business English: A Corpus-based Analysis. *English for Specific Purposes*. 25, 217–34
- O’Keeffe, A., McCarthy, M. & Carter, R. (2007). *From Corpus to Classroom*. Language Use and Language Teaching. Cambridge: Cambridge University Press.
- Oxford Collocations Dictionary* (2009). Oxford: Oxford University Press
- Partington, A. (1998). *Patterns and Meanings. Using Corpora of English Language Research and Teaching*. Amsterdam: 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.
- Phoocharoensil, S. (2010). A Corpus-based Study of English Synonyms. *International Journal of Arts and Sciences*. 3(10), 227-245.
- Phoocharoensil, S. (2020). English Get-passives: Reassessing the Frequencies across Genres. *GEMA Online® Journal of Language Studies*, 20(2), 123-135.
- Ruehleemann, C. (2010). What Can a Corpus Tell Us about Pragmatics? In O’Keefe A. & McCarthy, M. (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 288-302). London: Routledge.
- Schmitt, N. (2010). *Research Vocabulary. A Vocabulary Research Manual*. London: Palgrave Macmillan.
- Sinclair, J. (1999). *Corpus, Concordance, Collocation: Describing English Language*. Oxford: Oxford University Press.
- Sinclair, J. (2004). *Trust the Text*. London: Routledge.
- Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell
- Szudarski, P. (2018). *Corpus Linguistics for Vocabulary: A Guide for Research*. London: Routledge.
- Thornbury, S. (2002). *How to Teach Vocabulary*. Harlow: Longman
- Timmis, I. (2015). *Corpus Linguistics for ELT*. London: Routledge.
- Tognini-Bonelli, E. (1993). Interpretative Nodes in Discourse – *Actual and Actually*. In M. Baker, G. Francis & (Eds), *Text and Technology* (pp. 193–212). Amsterdam: John Benjamins.
- Wongkhan, P. & Thienthong, A. (2020). EFL Learners’ Acquisition of Academic Collocation and Synonymy: Does their Academic Experience Matter? *RELC Journal*. 1-16
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Xiao, R. & McEnery, T. (2006). Collocation, Semantic Prosody and Near Synonymy: A Cross-linguistic Perspective. *Applied Linguistics*. 27(1), 103-129.

ABOUT THE AUTHOR

Supakorn Phoocharoensil is an Associate Professor of English at Language Institute of Thammasat University, Bangkok, Thailand. He is also the Editor-in-Chief of LEARN Journal: Language Education and Acquisition Research Network, indexed in SCOPUS and ERIC. His areas of research interest include Second Language Acquisition, Corpus Linguistics, and English Collocations and Formulaic Language.