DIFFERENTIATING THE NEAR-SYNONYMS USING A CORPUS-BASED APPROACH: *BOOKING* VS *RESERVATION*

Piyapong Laosrirattanachai[a]; Supakorn Phoocharoensil[b]
([a]piyapong.l@ku.th; [b]supakorn.p@litu.tu.ac.th)

[a]Faculty of Hospitality Industry, Kasetsart University
1Moo 6, Kamphaeng Saen, Nakhonpathom 73140 Thailand

[b]Language Institute, Thammasat University
2 Phra Chan Alley, Phra Borom Maha Ratchawang, Phra Nakhon, Bangkok 10200 Thailand

Abstract: Precise use of near-synonymous vocabulary is essential for effective communication in the service industry, where subtle lexical distinctions can influence clarity, professionalism, and pragmatic appropriateness across different contexts. This study examined the corpus data pertaining to the synonymous terms *booking* and *reservation* to determine whether they are interchangeable in all contexts. The study explores variations in formality, genre distribution, collocations, grammatical patterns, semantic preferences, and semantic prosody. Data were sourced from the Corpus of Contemporary American English (COCA). The findings indicate that *reservation* is more formal, as it has higher Words Per Million values in formal genre. When examining collocates with an MI Score ≥3 and a frequency of occurrence of five or more, it was found that these two words shared only 12 out of 110 collocates. However, the two near-synonyms share all semantic preferences but have differing proportions of collocations covering each semantic field, particularly in some fields, namely SERVICE, IT AND COMPUTING, QUANTITY, and CERTAINTY. In terms of semantic prosody and colligation, the two near-synonyms show no significant differences, but it is interesting to note that the grammatical pattern "*make* + *reservation*" is used to a greater extent than "*make* + *booking*. It is recommended that L2 learners of English be familiarized with these similarities and differences of both synonyms to use English naturally and effectively in both general and industry-specific contexts within the service industry.

Keywords: collocation, corpus, degree of formality, near-synonyms, semantic preference

DOI: http://dx.doi.org/10.15639/teflinjournal.v36i2/350-371

English has long maintained considerable importance as a medium of communication owing to its widespread usage both locally and internationally, thereby establishing it as one of the most widely used languages globally (Rungrueang et al., 2022). A crucial element of effective communication is vocabulary, as it allows speakers to articulate their intended messages clearly. Conversely, a deficiency in vocabulary can impede listeners' comprehension of the speaker's message. This underscores the vital role of vocabulary in communication. Among vocabulary learning issues, synonyms are an essential element of language, aiding language learners in developing a deeper understanding of English and enhancing their effective communication (Boontam & Phoocharoensil, 2022; Pellicer-Sánchez et al., 2022; Webb et al.,

2012). However, synonyms in English pose a challenge for L2 learners, who often struggle to differentiate between words and their appropriate usage due to their similar meanings. Words with close meanings do not necessarily have identical usage (Crawford & Csomay, 2016). As a result, many second language (L2) learners are unable to use synonyms for communication in various contexts as appropriately as first language (L1) users (Yang et al., 2020). Therefore, enhancing the understanding of synonym usage for L2 learners, including the similarities, differences, and appropriateness in various contexts, is of utmost importance. This thorough understanding helps learners use synonyms more effectively (Edmonds & Hirst, 2002).

The terms *booking* and *reservation* are frequently used in service industries, such as tourism, hotels, and airlines. This is evident from the inclusion of both terms in word lists for the tourism, hotel, and airline businesses (Laosrirattanachai & Ruangjaroon, 2021a). Additionally, both terms are commonly found in everyday contexts, as demonstrated by their inclusion in the General Service List (West, 1953). Given their frequent use in both general and specialized contexts and their similar meanings, these terms may pose challenges for appropriate usage among students in specialized service industry fields as well as for those needing to use these terms in everyday communication. To differentiate between these terms, the simplest method is typically to consult dictionary definitions. Therefore, the authors have compiled information on the meanings of both terms from three online dictionaries, as shown in Table 1.

Table 1. Meanings of *booking* and *reservation* from Online Dictionaries

Words	**Meaning**		
	Cambridge	**Longman**	**Oxford**
Booking	an **arrangement** you make to have a hotel room, tickets, etc. at a particular time in the future, or the process of making this arrangement	an **arrangement** to travel by train, use a hotel room etc at a particular time in the future	an **arrangement** that you make in advance to buy a ticket to travel somewhere, go to the theatre, etc.
Reservation	an **arrangement** in which something such as a seat on an aircraft or a table at a restaurant is kept for you	an **arrangement** which you make so that a place in a hotel, restaurant, plane etc is kept for you at a particular time in the future	an **arrangement** for a seat on a plane or train, a room in a hotel, etc. to be kept for you

Table 1 demonstrates that the three dictionaries provide remarkably similar definitions for both *booking* and *reservation*, with the key meaning being an arrangement for the future use of something. This conclusion highlights a significant limitation of dictionaries: they offer only a rudimentary comprehension of these terms and fail to supply the detailed information essential for language learners. As a result, this limitation prevents learners from discerning the distinctions between these words and utilizing them inappropriately within various contexts (Ly & Jung, 2015). This similarity suggests that *booking* and *reservation* could be possibly considered “synonyms”.

Synonyms

The term synonym refers to words or phrases that have the same or very similar meanings to other words, though there may be some differences in certain aspects (Cruse, 2000; Taylor, 2002). In the case of the English language, which has been influenced by many other languages worldwide, English contains a large number of synonyms, both from its own development and from the influence of other languages (Palmer, 1997). Synonyms can be broadly categorized into two types. The first type is absolute synonyms (also called strict, perfect, or real synonyms), which are words that have almost identical meanings and can be used interchangeably in all contexts (Jackson & Amvela, 2007; Taylor, 2002). However, in English, absolute synonyms are rare or may not exist at all, as having many absolute synonyms could lead to language redundancy (Phoocharoensil, 2020a). The second type is near-synonyms (also called loose synonyms), which are words that have similar or close meanings but cannot be used interchangeably in all contexts (Cruse, 1986; Harley, 2006; Inkpen & Hirst, 2006). Furthermore, numerous studies have shown that the use of these synonyms in communication differs in several aspects, such as their formality level, senses of meaning, collocations, colligation, semantic preference, and semantic prosody (DiMarco et al., 1993). Therefore, a key consideration when encountering synonyms is whether the synonyms are absolute synonyms or near-synonyms.

In the field of vocabulary studies, lexical level and CEFR level are often used to differentiate vocabulary. The authors employ the VocabProfile program (Cobb, n.d.) to determine the lexical level of *booking* and *reservation* and find that *booking* falls within the first 4,000 high-frequency words, while *reservation* is among the first 2,000 high-frequency words. Considering the lexical frequency categories outlined by Schmitt and Schmitt (2014), which include high-frequency (1-3,000), mid-frequency (3,001-9,000), and low-frequency (9,001 and beyond), it can be concluded that *booking* is a mid-frequency word, whereas *reservation* is a high-frequency word. Additionally, according to the Oxford Learner's Dictionaries, the CEFR level of the two words indicates that *booking* is at level B2, while *reservation* is at level B1. These findings signal that *booking* and *reservation* are near-synonyms. Although dictionaries define both words with similar meanings, their lexical level, lexical frequency, and CEFR level differ, reflecting different usage contexts. This underscores the importance and challenge of research aimed at differentiating the two near-synonyms, *booking* and *reservation*, through a corpus-based approach.

Researching semantically related words is a significant undertaking in lexicography and linguistics, where corpus linguistics proves to be an invaluable tool. It is a method designed for an in-depth exploration of language (Kennedy, 1998; Lindquist & Levin, 2018). This approach utilizes diverse types of language data, such as authentic examples from native speakers (Aroonmanakun, 2011). Corpus linguistics offers essential quantitative and qualitative data for linguistic studies (O'Keeffe & McCarthy, 2010), encompassing elements such as word pairings, sentence structures, genres, language varieties, and word frequency (Moon, 2010). Therefore, when studying semantically related words, it is crucial to employ a corpus-based approach for studying synonyms rather than solely relying on dictionaries, as it reveals insights that dictionaries alone cannot fully capture (Gu, 2017; Li, 2019; Petcharat & Phoocharoensil, 2017). According to a number of studies on synonym analysis, the corpus-based approach is

instrumental in distinguishing near-synonyms in terms of genres, formality, collocation, and other aspects such as semantic preference, semantic prosody, and colligation.

Degrees of Formality

Discussions regarding the formality of language often revolve around a spectrum that spans from formal to informal language, positioned at opposite poles (Laosrirattanachai & Laosrirattanachai, 2024). Formal language is distinguished by its precision, coherence, and clarity, irrespective of context or shared knowledge. Conversely, informal language is typically more conversational, personalized, and context-dependent, relying heavily on mutual understanding and familiarity (Graesser et al., 2014). Despite the fact that synonyms may share similar meanings, their level of formality can vary depending on the context. Certain words may be more commonly found in academic literature, while others may be prevalent in everyday discourse (Jackson & Amvela, 2007). Among a set of synonymous terms, a combination of formal and informal usage may be observed, indicating variation in the level of formality among these words (Cruse, 1986; Phoocharoensil, 2020b). For example, within a group of three nearly synonymous words, *disadvantage* tends to be employed in formal settings, whereas *downside* and *drawback* are frequently used in informal contexts (Sumonsriworakun, 2022).

Collocation and Colligation

The examination of synonymous words involves a crucial consideration of collocation, referring to the natural pairing of two or more words forming associations (Baker et al., 2006; McCarthy & O'Dell, 2005; Timmis, 2015). For instance, within the domain of the hotel industry, the term *spacious room* exemplifies a collocation, wherein *spacious* and *room* naturally co-occur and are commonly used by native hotel personnel. In contrast, *big room* is less prevalent and not favored as a collocation. Determining collocations often relies on frequency values and Mutual Information (MI) Scores. Collocations with an MI Score below 3 are considered non-strong (Imsa-ard & Phoocharoensil, 2022). The study of collocations proves advantageous in discerning disparities among synonyms. Despite sharing akin meanings, divergent collocations may restrict their interchangeable usage (Phoocharoensil, 2025; Xiao & McEnery, 2006). For instance, although *small* and *little* are near-synonyms and share certain collocates, *small* tends to associate more frequently with terms such as *proportion*, *intestine*, and *firms*, whereas *little* exhibits a preference for terms like *while*, *evidence*, and *bit* over *small* (Aroonmanakun & Aroonmanakun, 2023).

Exploring near-synonyms extends to the examination of colligational patterns as an alternative approach. Despite conveying similar meanings, synonyms often lack complete interchangeability. This arises from the fact that near-synonyms articulate the same concept through diverse avenues and within varying contexts, each offering a distinct perspective. Consequently, their usage should be judiciously tailored to specific circumstances (Edmonds & Hirst, 2002). Colligation, or grammatical patterns, bears close resemblance to collocation but diverges in terms of word usage structure. Colligation relies on syntagmatic patterns and may encompass factors such as word order, tense, voice, or specific vocabulary within a given context (Flowerdew, 2012). For instance, in comparing the colligational patterns of *foresee* and *predict*,

there is a higher incidence of "*foresee + somebody/something + V.ing*" pattern for *foresee*, while *predict* tends to occur more frequently in the "*predict + wh-word*" pattern (Phoocharoensil, 2021a).

Semantic Preference and Semantic Prosody

Semantic preference refers to the inclination of lexical items that can be classified into groups based on specific semantic contexts and are part of different lexical sets (Begagić, 2013; Cheng, 2012; Hunston, 2002). In the differentiation of synonymous words, it is customary to examine semantic preference alongside collocation (Szudarski, 2018). For example, Kruawong and Phoocharoensil (2022) offer an illustration of the distinct semantic preferences among the synonymous verbs *teach*, *educate*, and *instruct*. Their research unveils that *teach* is predominantly utilized within the semantic domain of education and training, *educate* finds frequent usage in the realms of social work and family, and *instruct* is commonly employed in the context of laws and politics. However, Laosrirattanachai and Laosrirattanachai (2025) argue that although semantic preference is a valuable analytical tool for distinguishing near-synonyms, such distinctions should not rely on semantic preference alone. Near-synonymous items may exhibit substantial overlap in their semantic preferences despite underlying differences. For this reason, semantic preference should be examined in conjunction with collocational preference. Analyzing differences in collocational patterns within the same semantic preference category allows for a more fine-grained and nuanced differentiation between near-synonymous words.

Semantic prosody involves analyzing the patterns of word or phrase usage to ascertain whether the connotations of a word are positive, negative, or neutral, based on its collocates (Louw, 1993; Wynne, 2005). The core attributes of semantic prosody encompass three key factors (Hunston, 2002). First, the semantic prosody of a set of vocabulary pertains to the meaning derived from considering groups of words, rather than individual terms. Second, it reflects the implicit connotations of words that appear in specific contexts. Indicating whether a word has a positive, negative, or neutral meaning cannot be determined by analyzing the word in isolation but must be considered alongside its collocating words. (Partington, 1998). Third, the reliable assessment of a word's semantic prosody must be based on the examination of large language datasets, such as corpora, which provide credible information and reflect the word's usage in a wide range of authentic communicative contexts (Hunston, 2002; Szudarski, 2018). Therefore, when evaluating the semantic prosody of near-synonyms, it is essential to adhere to these three criteria. This involves not only considering the words within the near-synonyms themselves but also examining the associated collocations. Additionally, such analysis should be conducted using extensive linguistic data sources, such as major corpora like the British National Corpus (BNC) or the Corpus of Contemporary American English (COCA). For example, an analysis of the semantic prosody of the near-synonyms *persist* and *persevere* indicates that *persist* tends to co-occur more frequently with words or phrases that convey negative meanings (Phoocharoensil, 2021b). However, although near-synonyms cannot be used interchangeably in an absolute sense and various analytical approaches can reveal their differences, not all sets of near-synonyms differ across every dimension. Rather, near-synonyms may exhibit distinctions in certain aspects, such as collocational behavior, while remaining

largely interchangeable with respect to other dimensions, including degree of formality and semantic prosody (Sridhanyarat & Phoocharoensil, 2023).

The present investigation seeks to scrutinize the synonymous terms *booking* and *reservation* concerning their distribution across genres, formality levels, collocational patterns, semantic preferences, semantic prosody, and colligations, utilizing the COCA—a substantial English corpus comprising over one billion words. The findings of this inquiry hold the potential to enrich the comprehension of L2 learners regarding the resemblances and disparities between the two words. This corpus-based approach provides a clearer picture of how *booking* and *reservation* are used, underscoring the importance for L2 learners to explore these synonyms through corpus data in tandem with traditional dictionaries to attain a more comprehensive mastery of their usage.

Research Questions

The current study seeks to address the following research questions:

1. What are the similarities and distinctions between the near-synonyms *booking* and *reservation* in terms of genre distribution and degree of formality?
2. How do the collocations, semantic preferences, and semantic prosody of the near-synonyms *booking* and *reservation* differ?
3. What are the similarities and differences between the near-synonyms *booking* and *reservation* in terms of colligations?

METHOD

The present study examined two synonymous nouns, *booking* and *reservation*. In the preceding section, it was demonstrated that *booking* and *reservation* are considered near-synonyms based on their lexical levels, frequency of use, and CEFR levels. Consequently, this study sought to delineate the differences between these near-synonyms concerning genre distribution, levels of formality, collocational patterns, semantic preferences, semantic prosody, and colligations. The data used in this research was sourced from the COCA, which is well-known as a comprehensive corpus of American English, containing over one billion words across eight genres, including spoken language, fiction, magazines, newspapers, academic texts, web pages, blogs, and TV and movie scripts (Davies, 2021). The COCA has been widely acknowledged by researchers and scholars globally as a reliable source of data for conducting language-related research (Liu, 2010; Yoo & Shin, 2020).

Data Analysis

The data analysis process encompasses the following steps.

Analyzing Genre Distribution and Formality Levels

The SEARCH function was utilized with the Chart command being selected to observe the frequency and Words Per Million (WPM) values of words within the eight genres. Given that

this research focuses on *booking* and *reservation*, both of which function as nouns, it is necessary to specify to display only nouns, such as "booking [NOUN]." Subsequently, the two near-synonyms were analyzed by prioritizing WPM values as the primary metric. Since each genre varies in size, assessing solely based on raw frequency may lead to bias, hence the necessity to consider WPM values instead. If a near-synonym appears in any genre with a significantly higher WPM value than the other one, it can be concluded that the former near-synonym is more commonly used in that genre than the latter. In the assessment of formality levels, the WPM values were also taken into account. TV and Movie subtitles and spoken transcripts were categorized as informal, while academic articles were considered formal. The remaining five genres fell somewhere in between these two extremes (Davies, 2021). Therefore, when considering the formality of near-synonyms, it was necessary to evaluate the WPM values resulting from the aggregation of genres within the same formality level.

Analyzing Collocations

In considering collocation, the COLLOCATE function was used to find words that appear together with the target near-synonyms. There are two methods for analyzing collocation: examining the frequency value and the MI score (Schmitt, 2010). The MI score identifies the strength of the relationship between two words when they appear together, taking into account their co-occurrence frequency with all other words in the corpus (Hunston, 2002). However, even if a particular collocation has a high MI score, it may not be widely used by language users, as indicated by its low raw frequency in language corpora. This suggests that it is not actually employed extensively (Szudarski, 2018). Therefore, to balance the bias of using solely frequency or MI score, both values are considered. Accordingly, this study utilized an MI score threshold of ≥ 3. For the frequency cut-off point, it was set at 5 occurrences or more (Nesselhauf, 2005).

Analyzing Semantic Preferences and Semantic Prosody

After obtaining the set of collocations, the words that meet the frequency and MI score criteria were considered for their semantic preference and semantic prosody. To facilitate the analysis of both semantic preference and semantic prosody, the authors proposed using the UCREL Semantic Analysis System (USAS) (Piao et al., 2015). The USAS is an online program that helps categorize words into 21 major semantic fields (see https://ucrel.lancs.ac.uk/usas/). Additionally, the USAS can identify the semantic prosody of words, determining whether a word has a positive or negative connotation. However, some words may be categorized into more than one semantic group due to the prevalence of polysemous words in English. Therefore, the USAS was used in the initial stage to analyze the entire set of collocations. The authors then manually refined the analysis by rechecking and selecting words with multiple results to ensure the most reliable outcomes. For example, the collocation *gross booking* is used to refer to the total value of all reservations made through a travel service or platform before deducting cancellations, refunds, and fees. However, the word *gross* can mean "*total*" or "*extremely unpleasant*." Thus, the analysis results from the USAS are as follows:

gross = N5.1+, O4.2-
booking = A9+

From the displayed categorization, the word *gross* can be categorized under N5.1+, which belongs to the semantic field related to "*entirety or maximum*" and carries a positive aura. Conversely, when categorized under O4.2-, it falls into the semantic field related to "*judgement of appearance*" and carries a negative aura. This necessitated a manual analysis by the authors. As a result, N5.1+ was selected as it aligns with the contextual meaning. Meanwhile, *booking* is categorized under A9+, which pertains to the semantic field of "*getting and giving or possession*" and is already appropriate, thus not requiring a manual recheck.

Analyzing Grammatical Patterns

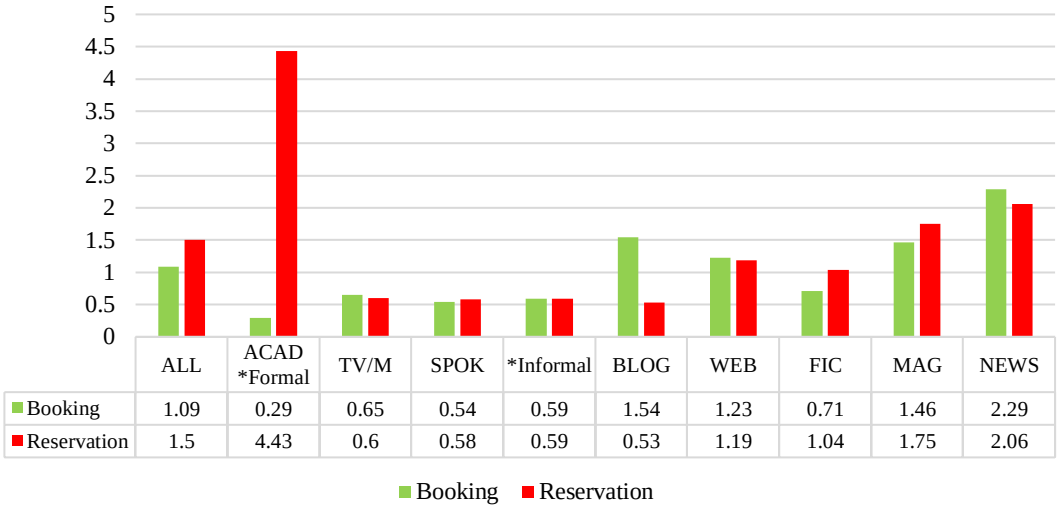
The Key Word in Context (KWIC) tool was used to extract the colligations of the near-synonyms *booking* and *reservation*. To conduct the analysis, the first 200 concordance lines for each near-synonym were examined. Concordance lines with the same patterns were grouped, and their frequency was counted to identify and summarize the prominent and commonly used colligations for each target near-synonym.

FINDINGS AND DISCUSSION

To address the research questions and illustrate the similarities and differences of the near-synonyms *booking* and *reservation*, the COCA was used to identify the following issues.

Genre Distribution and Degree of Formality

To address Research Question 1, "*What are the similarities and distinctions between the near-synonyms booking and reservation in terms of genre distribution and degree of formality?*", Figure 1 displays the Words Per Million (WPM) values of the near-synonyms *booking* and *reservation* across eight genres. Overall, *reservation* appeared at a rate of 1.5 occurrences PM words, whereas *booking* appeared at a rate of 1.09 occurrences PM words. This suggests that, in general, there is not a significant difference in the frequency of use between these two near-synonyms in English. However, the WPM values of *reservation* (1.5) and *booking* (1.09) in COCA align with their lexical levels and CEFR levels. Specifically, the higher WPM value of *reservation* corresponds to its classification as a high-frequency word, appearing within the first 2,000 high-frequency words and is at the B1 level of the CEFR. In contrast, *booking*, with a lower WPM value, is classified as a mid-frequency word, appearing within the first 4,000 high-frequency words and is at the B2 level of the CEFR.



	ALL	ACAD *Formal	TV/M	SPOK	*Informal	BLOG	WEB	FIC	MAG	NEWS
Booking	1.09	0.29	0.65	0.54	0.59	1.54	1.23	0.71	1.46	2.29
Reservation	1.5	4.43	0.6	0.58	0.59	0.53	1.19	1.04	1.75	2.06

Figure 1. Genre Distribution and Degree of Formality of *booking* and *reservation*

When considering genre distribution, the analysis shows that across eight genres, the two synonymous nouns exhibit very similar WPM values in six of them, namely spoken language, fiction, magazines, newspapers, web pages, and TV and movie scripts. This pattern suggests that in these genres, *booking* and *reservation* are largely interchangeable and tend to be used with comparable frequency. However, notable differences emerge in the remaining two genres, indicating that genre-specific communicative norms influence lexical choice. In academic texts, *reservation* displays a substantially higher WPM value than booking (4.43:0.29), reflecting its association with formality and conventional academic style. In contrast, in blogs—commonly regarded as a form of informal language—*booking* occurs more frequently than *reservation* (1.54:0.53). This contrast can be explained by the communicative characteristics of blogs, which typically favor a more conversational, personalized style and rely heavily on shared background knowledge between writers and readers (Graesser et al., 2014). Such features make *booking* a more natural lexical choice in this genre, as blogs often function as platforms for individuals to express personal experiences and opinions to audiences who, despite their diversity, share sufficient mutual understanding. Taken together, these findings indicate that *booking* is more frequently used in blogs, *reservation* is preferred in academic texts, and in the remaining genres, both terms tend to occur with similar frequency.

Regarding the degree of formality, it was found that in formal contexts, *reservation* is the preferred term. This is evidenced by the significantly higher WPM value of *reservation* in the academic genre, which is categorized as formal. In contrast, in informal contexts, which include the spoken language and TV and movie scripts genres, the WPM values of *booking* and *reservation* are nearly identical. These observations support the claim that lexical choices are sensitive to context, with certain words occurring more frequently in academic or formal discourse, while others are more prevalent in everyday language (Jackson & Amvela, 2007). In

light of this distinction, it can be concluded that L2 learners should preferentially select *reservation* when communicating in formal or academic contexts, where more conventionalized and formal lexical choices are expected.

In the following section, Table 2, Table 3, and Figure 2 address Research Question 2: "*How do the collocations, semantic preferences, and semantic prosody of the near-synonyms booking and reservation differ?*"

Collocations

In this study, the noun, verb, and adjective collocations of *booking* and *reservation* were compared. The criteria for collocations to be considered were that they must co-occur with the target near-synonyms with an MI score of ≥ 3 and a frequency of 5 occurrences or more. Additionally, collocations classified as proper names, as well as those irrelevant to the context of "*an arrangement for the future use of something*," were not considered. For instance, *sabre* was excluded from the analysis because, when it co-occurs with *reservation*, it constitutes the proper name *'Sabre Reservation*', which is the name of a travel reservation system. The noun, verb, and adjective collocations of *booking* and *reservation* that met both conditions are shown in Table 2.

Table 2. Noun, Verb, and Adjective Collocations of *booking* and *reservation*

Collocations	Booking	Reservation	Shared
Noun	agency, agent, app, award, band, block, concert, convention, cruise, desk, discount, drop, e-mail, engine, entertainment, fare, fee, flight, golf, guest, handler, holiday, passenger, request, resort, revenue, sheet, site, slip, theater, vacation, website	allotment, bar, beer, campground, camping, campsite, cancellation, canyon, card, dining, dinner, entrée, ferry, gratuity, inn, lodging, motel, nonmember, occupancy, parking, quota, restaurant, seating, standing, superior, tour, valet, visa, walk-in, weekday, wheelchair, wine	advance, airline, clerk, hotel, lunch, rental, ticket, travel
Verb	advance, arrange, handle, process, secure, travel	accept, allot, cash, lodge, recommend, request, require, service, snag	cancel, confirm
Adjective	advanced, corporate, double, future, gross, online, tee-time, undated	accepted, canceled, casual, computerized, confirmed, guaranteed, guided, makeshift, on-site, third-party, toll-free	last-minute, same-day
Total 110 (100%)	**46 (41.82%)**	**52 (47.27%)**	**12 (10.91%)**

Table 2 illustrates that although *booking* and *reservation* are near-synonyms with similar dictionary definitions, data from the COCA shows that they share very few collocations when used with other collocating words. It is evident that out of the 110 collocations meeting the recommended criteria of frequency and mutual information score (Imsa-ard & Phoocharoensil,

2022; Nesselhauf, 2005), and aligning with the contextual meanings of *booking* and *reservation* as defined in this study, only 12 collocations (10.91%) are shared by both terms. This clearly reflects that in terms of collocation, *booking* and *reservation* differ significantly and cannot be used interchangeably in all contexts (Harley, 2006; Inkpen & Hirst, 2006). Each near-synonym is suitable for use with different collocations, as illustrated in the following examples.

(1) flight ***booking/*reservation***

(2) restaurant ***booking/reservation***

(3) holiday ***booking/*reservation***

(4) standing ***booking/reservation***

(5) gross ***booking/*reservation***

(6) on-site ***booking/reservation***

Examples (1)-(6) illustrate that the near-synonyms *booking* and *reservation* have different suitable collocations. For instance, when referring to booking a flight, *flight booking* is preferred over **flight reservation*. Similarly, when referring to booking something on-site, *on-site reservation* is preferred over **on-site booking*.

In communication, while L1 and L2 users of English can still understand each other even if they do not use the appropriate collocations, emphasizing the understanding of proper collocations is crucial (Yang et al., 2020). It leads to smoother, more natural, and more effective communication with native speakers. For instance, consider the collocation "*(2) restaurant *booking/reservation.*" When examining the collocations that can occur, both *restaurant booking* and *restaurant reservation* are grammatically correct and convey the same intended meaning. However, although the data from the COCA indicates that *restaurant booking* is also in use, *restaurant reservation* is higher in frequency and MI score, highlighting the stronger relationship between *restaurant* and *reservation* when they appear together (Xiao & McEnery, 2006). Given that the primary goal of teaching collocations is to help L2 learners acquire language that sounds more natural, it is essential to stress this understanding to learners. This will enable them to communicate naturally and effectively in relevant contexts, closely approximating native speaker proficiency.

Semantic Preference and Semantic Prosody

After analyzing the semantic preference and semantic prosody of the near-synonyms *booking* and *reservation* by first using the USAS and then manually cross-checking, the findings are illustrated in Table 3 and Figure 2.

Table 3 shows that the collocations of *booking* and *reservation* can be categorized according to their semantic preferences into 13 semantic fields, based on the USAS system and manual cross-checking. Overall, *booking* and *reservation* can be used within the same semantic fields, but with different collocations. For instance, both *booking* and *reservation* can be used in the semantic field of IT AND COMPUTING. However, *booking* is collocated with terms such as

app, *e-mail*, *engine*, *online*, and *website*, whereas *reservation* is collocated with *computerized*. This divergence in collocational patterns provides a valuable basis for distinguishing between near-synonymous lexical items (Phoocharoensil, 2025).

Table 3. Semantic Preference of Collocations of *booking* and *reservation*

	Semantic field	**Near-synonym**		
		Booking	**Shared**	**Reservation**
1	BUSINESS 19 (100%)	(+0) (±8) agency, concert, convention, corporate, cruise, golf, resort, theater (–0)	(+0) (±3) airline, hotel, rental (–0)	(+0) (±8) bar, camping, ferry, inn, lodge, motel, restaurant, tour (–0)
		8 (42.11%)	**3 (15.78%)**	**8 (42.11%)**
2	CERTAINTY 6 (100%)	(+0) (±0) (–0)	(+1) confirm (±0) (–0)	(+4) accept, accepted, confirmed, guaranteed (±1) casual (–0)
		0 (0.00%)	**1 (16.66%)**	**5 (83.34%)**
3	HELPING/ HINDERING 6 (100%)	(+2) award, sponsorship (±0) (–0)	(+0) (±0) (–1) cancel	(+1) snag (±0) (–2) canceled, cancellation,
		2 (33.33%)	**1 (16.67%)**	**3 (50.00%)**
4	IT AND COMPUTING 6 (100%)	(+0) (±5) app, e-mail, engine, online, website (–0)	(+0) (±0) (–0)	(+0) (±1) computerized (–0)
		5 (83.34%)	**0 (0.00%)**	**1 (16.66%)**
5	MONEY 8 (100%)	(+1) revenue (±0) (–3) discount, fare, fee	(+0) (±0) (–0)	(+2) cash, gratuity (±2) toll-free (–0)
		4 (50.00%)	**0 (0.00%)**	**4 (50.00%)**
6	OBJECT 7 (100%)	(+0) (±3) photo, sheet, slip (–0)	(+0) (±1) ticket (–0)	(+0) (±3) card, visa, wheelchair (–0)
		3 (42.86%)	**1 (14.28%)**	**3 (42.86%)**
7	ORGANIZING 6 (100%)	(+2) arrange, process (±0) (–2) handle, secure	(+0) (±0) (–0)	(+2) allot, allotment (±0) (–0)
		4 (66.67%)	**0 (0.00%)**	**2 (33.33%)**
8	PEOPLE 8 (100%)	(+0) (±5) agent, band, guest, handler, passenger	(+0) (±1) clerk (–0)	(+0) (±2) nonmember, third-party (–0)

	Semantic field	Near-synonym		
		Booking	**Shared**	**Reservation**
		(–0)		
		5 (62.50%)	**1 (12.50%)**	**2 (25.00%)**
9	PLACE 6 (100%)	(+0) (±2) desk, site (–0)	(+0) (±0) (–0)	(+0) (±4) campground, campsite, canyon, on-site (–0)
		2 (33.33%)	**0 (0.00%)**	**4 (66.67%)**
10	QUANTITY 5 (100%)	(+2) block, gross (±0) (–2) double, drop	(+0) (±0) (–0)	(+1) quota (±0) (–0)
		4 (80.00%)	**0 (0.00%)**	**1 (20.00%)**
11	SERVICE 17 (100%)	(+1) entertainment (±1) flight (–0)	(+0) (±2) lunch, travel (–0)	(+0) (±13) beer, dining, dinner, entrée, guided, lodging, occupancy, parking, seating, service, valet, walk-in, wine (–0)
		2 (11.77%)	**2 (11.77%)**	**13 (76.46%)**
12	SPEECH ACTS TERMS 3 (100%)	(+0) (±0) (–0)	(+0) (±0) (–1) request	(+1) recommend (±0) (–1) require
		0 (0.00%)	**1 (33.33%)**	**2 (66.67%)**
13	TIME 12 (100%)	(+5) advanced, future, tee-time, vacation, holiday (±0) (–1) undated	(+1) advance (±1) same-day (–1) last-minute	(+1) weekday (±1) standing (–1) makeshift
		6 (50.00%)	**3 (25.00%)**	**3 (25.00%)**

***Remark**: (+) = Positive, (±) = Neutral, (–) = Negative*

The analysis of the semantic preferences of *booking* and *reservation* reveals clear differences in four semantic fields. In the SERVICE field, *reservation* is used 76.46%, compared to only 11.77% for *booking*. Similarly, *reservation* is more frequently used in CERTAINTY (83.34%). On the other hand, in the fields of IT AND COMPUTING and QUANTITY, *booking* is more frequently used at proportions of 83.34% and 80.00%, respectively. The findings indicate that investigations of semantic preference should also take into account the collocational behavior of near-synonyms within the same semantic preference category. Such an approach makes it possible to demonstrate that, although near-synonyms may share broad semantic preferences, they nevertheless differ in their selection of collocations within those categories

(Laosrirattanachai & Laosrirattanachai, 2025). Examples (7) to (10) illustrate the use of the near-synonyms in the data:

(7) Things have settled down now, although ***parking reservations*** remain tight. [SERVICE]

(8) To be eligible for compensation though, you must have a ***confirmed reservation*** and arrive at the airport on time. [CERTAINTY]

(9) While most hotels tend to show up on every major ***booking website***, rental inventory varies from site to site. [IT AND COMPUTING]

(10) This is what Uber is doing, their -- what it has been looking at Uber Eats for example, ***gross bookings*** on ride hailing was up 22 percent, Uber Eats was up 108 percent. So it's much faster growing. [QUANTITY]

The analysis from Table 3 is then summarized to show the proportions of semantic prosody, as shown in Figure 2.

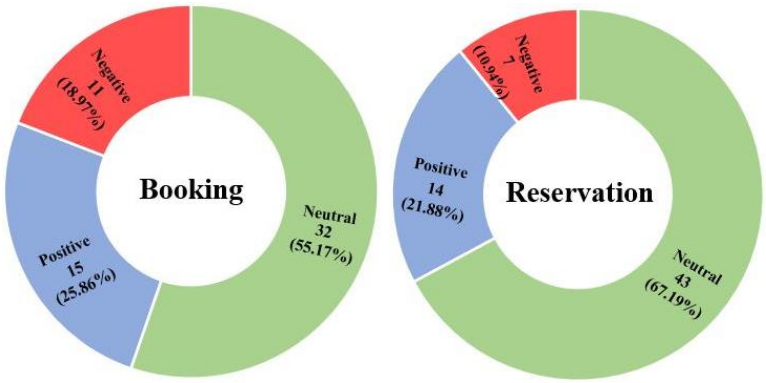


Figure 2. Semantic Prosody Proportions of Collocations of *booking* and *reservation*

Figure 2 demonstrates that the semantic prosody of *booking* and *reservation* tends towards a neutral sentiment, covering 55.17% and 67.19%, respectively. This indicates that communication with a positive or negative inclination does not significantly influence the choice between *booking* and *reservation*. It should be noted that, having concluded that *booking* and *reservation* are near-synonyms, this means these two words have similar meanings and can be used interchangeably only in certain contexts (Edmonds & Hirst, 2002). In other contexts, they cannot be used interchangeably (Cruse, 1986; Harley, 2006; Inkpen & Hirst, 2006), as evidenced by previous analyses regarding genre distribution, degree of formality, and collocation. Therefore, the neutrality in semantic prosody for both *booking* and *reservation* suggests that these terms can be used interchangeably in various contexts without the risk of conveying unintended positive or negative connotations (Sridhanyarat & Phoocharoensil, 2023). This neutrality is advantageous for users, as it provides flexibility in language use without the concern

of altering the tone of the communication, especially in contexts where *booking* and *reservation* can be used interchangeably. Examples (11) to (14) illustrate the semantic prosody of the near-synonyms in the data:

Booking

(11) Crucial to the B marketplace, ***block booking*** enabled distributors to package films in groups. (Positive)

(12) You are a girl-about-town, but the only danger is ***double booking***. (Negative)

Reservation

(13) By luck, I ***snagged a reservation*** at one of the toughest tables in town to get. (Positive)

(14) If your travel authorization includes a domestic flight, it must be approved and ticketed at least 72 hours in advance of your departure to avoid a ***reservation cancellation***. (Negative)

Colligations

To address Research Question 3, "*What are the similarities and differences between the near-synonyms booking and reservation in terms of colligations*?", the concordance lines in the COCA were analyzed. After analyzing the first 300 concordance lines of each target near-synonym, it was found that both *booking* and *reservation* exhibit a wide variety of grammatical patterns. Both terms appear in all forms that English nouns can take, leading to the conclusion that there are no significant differences in grammatical patterns between *booking* and *reservation* in terms of colligation, which is aligned with Sridhanyarat and Phoocharoensil (2023). However, an interesting pattern emerged from the data: English speakers tend to use the verb *make* with *reservation* significantly more often. In line with a study by Laosrirattanachai and Laosrirattanachai (2025), this finding underscores the importance of examining collocation when distinguishing between near-synonyms. If analysis is limited to colligation alone, such differences may remain obscured, since the colligational pattern *verb + booking/reservation* is highly common.

A closer collocational analysis, however, reveals that the verb *make* shows a clear preference for co-occurring with *reservation*. Examination of the COCA data revealed 375 instances of the grammatical pattern "*make + reservation*," compared to only 15 instances of "*make + booking*." Since the MI score of the collocation *make reservation* did not meet the specified threshold (the MI score of *make + reservation* is 2.32), *make* did not appear in the collocation analysis section. Examples (15) to (20) illustrate the similar colligational patterns shared by *booking* and *reservation*.

Colligational pattern 1 (verb + booking/reservation)

(15) … because I didn't ***have*** enough ***bookings*** in that month, and I thought something was wrong with me.

(16) We do ***have*** a ***reservation*** at Cinderella's for breakfast also so that will take up some…

Colligational pattern 2 (booking/reservation + preposition)

(17) He explained that 90 percent of the ***bookings for*** the event are made by him picking up the phone…

(18) We spoke on the phone a month ago to confirm our ***reservation for*** the house this weekend.

Colligational pattern 3 (noun + booking/reservation)

(19) We book hotel rooms through ***mediator booking*** companies.

(20) …can automatically start rebooking people inside the ***computer reservation*** systems quicker than we can talk to that human behind the gate…

CONCLUSION

Figure 3. Summary of Similarities and Differences of *booking* and *reservation*

The findings summarized in Figure 3 reveal that although *booking* and *reservation* have a similar core meaning, they are not universally interchangeable across all contexts. This is due to

their differences in aspects such as genre distribution, formality level, collocation patterns, and semantic preferences. Specifically, when examining eight genres, it becomes clear that *reservation* is predominantly used in academic texts, which are considered formal contexts, whereas *booking* is more commonly found in blogs. In terms of degree of formality, it was found that *booking* is used in both formal and informal contexts in roughly equal proportions. In contrast, the analysis clearly shows that *reservation* is more commonly used in formal contexts.

The use of collocation in analyzing the differences between *booking* and *reservation* yielded the most interesting and clear results compared to other methods. The analysis shows that *booking* and *reservation* share very few collocations—only 12 shared collocations, accounting for 10.91%. This indicates that *booking* and *reservation* cannot be used interchangeably when paired with different collocates. However, it is noteworthy that not choosing collocates with *booking* and *reservation* according to the current study's results does not imply grammatical incorrectness or cause communication breakdown. Instead, using the collocates with *booking* and *reservation* as shown in Table 2 enhances the ability to communicate more naturally, effectively, and closely to that of native speakers.

Regarding semantic preference, no differences were found between the target near-synonyms and the semantic fields in which they appear, as both *booking* and *reservation* can be used in the same semantic fields. However, there are two significant points to note. First, although both *booking* and *reservation* can be used in the same semantic fields, the collocates that appear with them in the same fields rarely overlap. Second, while *booking* and *reservation* generally appear in the same semantic fields, some semantic fields show a significant difference in the number of collocations for each term. For instance, in the semantic field of SERVICE, *reservation* has 13 collocations (76.46%), whereas *booking* has only 2 collocations (11.77%), and there are 2 shared collocations (11.77%).

In contrast to using collocation to differentiate between *booking* and *reservation*, the analysis results show that there was almost no difference between *booking* and *reservation* in terms of semantic prosody and colligation. Many scholars assert that while near-synonyms are interchangeable in certain contexts, they are not in others (Cruse, 1986; Harley, 2006; Inkpen & Hirst, 2006). Consequently, the near-synonyms *booking* and *reservation* are not interchangeable with respect to genre distribution, degree of formality, collocations, and semantic preferences. However, they can be used interchangeably in terms of semantic prosody and colligations.

Nonetheless, these findings possess certain limitations. Firstly, the data was sourced solely from the COCA, reflecting only American English. Collecting data from the BNC, which represents British English, could potentially provide further insights that differ from these results.

Pedagogical Implications

Relying solely on traditional dictionaries for studying near-synonyms may not yield comprehensive insights and fails to highlight the crucial distinctions needed by learners learning English as an additional language (Hunston, 2002). Therefore, teachers should emphasize the importance of distinguishing between synonymous words by teaching collocations to L2 learners, as collocations often pose significant challenges for learners at both basic and advanced

levels. Instruction in this area can help reduce errors in word usage across various contexts and enhance understanding of the nuanced differences among synonyms (Jafarpour et al., 2013).

The following recommendations are proposed for pedagogical implications: To deepen L2 learners' understanding of the near-synonyms *booking* and *reservation*, utilizing corpus data is essential for both teachers and students. Teachers can introduce learners to corpus data by examining concordance lines. For instance, they can use tools like the COCA or the BNC to show real-life examples of *booking* and *reservation* in context. This can help students understand how these near-synonyms are used differently depending on the context. This method is beneficial as it encourages students to explore and study independently, thereby improving comprehension (Boulton, 2011; Laosrirattanachai & Ruangjaroon, 2021b; Lee et al., 2020; Otto, 2021; Saeedakhtar et al., 2020). In class, teachers can provide lists of common collocations for each near-synonym. For example, pair *booking* with terms like *online*, *agent*, and *flight*, and *reservation* with *parking*, *dinner*, and *on-site*. Encourage students to use these lists in their writing and speaking exercises. Outside the classroom, teachers can develop interactive exercises or games using platforms like Quizlet or Kahoot to reinforce the correct usage of *booking* and *reservation* (Chen et al., 2019; Roohani & Heidari Vincheh, 2023). For example, they can create a matching game where students pair near-synonyms with their appropriate collocations. Additionally, teacher can utilize social media platforms or massive open online courses to engage students (Korucu-Kış, 2023; Martín-Monje & Borthwick, 2021). This may involve creating short instructional videos on TikTok or YouTube that explain the differences between near-synonyms. Then, share these videos along with exercises or quizzes to test understanding. These platforms have the potential to promote student engagement and enable the exchange of knowledge (Jin et al., 2022; Phoocharoensil, 2023; Yeh & Mitric, 2021). The present environment of online media offers convenient access and diverse avenues for interactive education and information dissemination (Loewen et al., 2019).

REFERENCES

Aroonmanakun, V., & Aroonmanakun, W. (2023). A comparison of 'little' and 'small' in English and Thai: a corpus-based study. *LEARN Journal: Language Education and Acquisition Research Network, 16*(1), 296–319. https://so04.tci-thaijo.org/index.php/LEARN/article/view/263443

Aroonmanakun, W. (2011). Corpora and emerging technology for ELT. *Journal of Studies in the English Language, 6*(1), 221–236. https://so04.tci-thaijo.org/index.php/jsel/article/view/21857

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics.* Edinburgh University Press.

Begagić, M. (2013). *Semantic preference and semantic prosody of the collocation make sense. Jezikoslovlje, 14*(2–3), 403–416. https://hrcak.srce.hr/112198

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

Boulton, A. (2011). Data-driven learning: the perpetual enigma. In S. Goźdź-Roszkowski (Ed.), *Explorations across languages and corpora* (pp. 563-580.). Peter Lang.

Cambridge Dictionary. (Online). https://dictionary.cambridge.org/

Chen, C.-M., Liu, H., & Huang, H.-B. (2019). Effects of a mobile game-based English vocabulary learning app on learners' perceptions and learning performance: A case study of Taiwanese EFL learners. *ReCALL, 31*(2), 170–188. https://doi.org/10.1017/S0958344018000228

Cheng, W. (2012). *Exploring corpus linguistics: Language in action.* Routledge.

Cobb, T. (n.d.). *Vocabprofile*. (Online programme). http://www.lextutor.ca/vp/

Crawford, W. J., & Csomay, E. (2016). *Doing corpus linguistics.* Routledge.

Cruse, D. A. (1986). *Lexical semantics.* Cambridge University Press.

Cruse, D. A. (2000). *Meaning in language: An introduction to Semantics and Pragmatics.* Oxford University Press.

Davies, M. (2021). *The new Corpus of Contemporary American English (COCA 2021)*. https://www.english-corpora.org/coca/

DiMarco, C., Hirst, G., & Stede, M. (1993). *The semantic and stylistic differentiation of synonyms and near-synonyms.* AAAI Spring Symposium on Building Lexicons for Machine Translation.

Edmonds, P., & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics, 28*(2), 105–144. https://doi.org/10.1162/089120102760173625

Flowerdew, L. (2012). *Corpora and Language Education.* Palgrave Macmillan.

Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). CohMetrix measures text characteristics at multiple levels of language and discourse. *The Elementary School Journal, 115*(2), 210–229. https://doi.org/10.1086/678293

Gu, B. J. (2017). Corpus-based study of two synonyms 'obtain' and 'gain'. *Sino-US English Teaching, 14*(8), 511–522. https://doi.org/10.17265/1539-8072/2017.08.006

Harley, H. (2006). *English words: A linguistic introduction*. Blackwell.

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

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

Inkpen, D., & Hirst, G. (2006). Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics*, *32*(2), 223–262. https://doi.org/10.1162/coli.2006.32.2.223

Jackson, H., & Amvela, E. Z. (2007). *Words, meaning and vocabulary: An introduction to modern English lexicology.* Bloomsbury.

Jafapour, A. A., Hashemian, M., & Alipour, S. (2013). A corpus-based approach toward teaching collocation of synonyms. *Theory and Practice in Language Studies, 3*(1), 51–60. https://doi.org/10.4304/tpls.3.1.51-60

Jin, H., Karatay, Y., Bordbarjavidi, F., Yang, J., Kochem, T., Muhammad, A. A., & Hegelheimer, V. (2022). Exploring global online course participants' interactions: Value of high-level engagement. *ReCALL, 34*(3), 291–308. doi:10.1017/S0958344021000331

Kennedy, G. (1998). *An introduction to corpus linguistics*. Longman.

Korucu-Kış, S. (2023). Instag(R)ite: Integrating visual social media into academic writing instruction. *Computer Assisted Language Learning*, 1–32. https://doi.org/10.1080/09588221.2023.2228838

Kruawong, T., & Phoocharoensil, S. (2022). A genre and collocational analysis of the near-synonyms teach, educate and instruct: A corpus-based approach. *TEFLIN Journal, 33*(1), 75–96. https://doi.org/10.15639/teflinjournal.v33i1/75-97

Laosrirattanachai, P. & Laosrirattanachai, P. (2024). The triangle of language use: A corpus-based analysis of hotel responses to reviews. *Teaching English as a Second Language Electronic Journal (TESL-EJ), 28*(1). https://doi.org/10.55593/ej.28109a3

Laosrirattanachai, P., & Laosrirattanachai, P. (2025). Unveiling the distinction of near synonymy: A Corpus-based analysis on *attempt*, *endeavor*, *strive*, and *try*. *PASAA, 70*, 132–163. https://doi.org/10.58837/CHULA.PASAA.70.5

Laosrirattanachai, P., & Ruangjaroon, S. (2021a). Corpus-based creation of tourism, hotel, and airline business word lists. *LEARN Journal: Language Education and Acquisition Research Network, 14*(1), 50–86. https://so04.tci-thaijo.org/index.php/LEARN/article/view/248677

Laosrirattanachai, P., & Ruangjaroon, S. (2021b). Implementation of a data-driven hospitality lexis learning programme. *3L: The Southeast Asian Journal of English Language Studies*, 27(1), 1–21. http://doi.org/10.17576/3L-2021-2701-01

Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the establishment of a data-driven learning model: role of learner factors in corpus-based second language vocabulary learning. *The Modern Language Journal, 104*(2), 345–362. https://doi.org/10.1111/modl.12634

Li, E. (2019). A corpus-assisted study of synonyms in EFL teaching: Take *preserve* and *conserve* as example. *Linguistics and Literature Studies, 7*(2), 39–50. https://doi.org/10.13189/lls.2019.070201

Lindquist, H. & Levin, M. (2018). *Corpus linguistics and the description of English* (2nd ed.). Edinburgh University Press.

Liu, D. (2010). Is It a *chief*, *main*, *major*, *primary*, or *principal* concern?: A corpus-based behavioral profile study of near-synonyms. *International Journal of Corpus Linguistics, 15*(1), 56–87. https://doi.org/10.1075/ijcl.15.1.03liu

Loewen, S., Crowther, D., Isbell, D. R., Kim, K. M., Maloney, J., Miller, Z. F., & Rawal, H. (2019). Mobile-assisted language learning: A Duolingo case study. *ReCALL, 31*(3), 293–311. https://doi.org/10.1017/S0958344019000065

Longman Dictionary of Contemporary English Online. (Online). https://www.ldoceonline.com/

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). John Benjamins.

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. http://journalarticle.ukm.my/9073/

Martín-Monje, E., & Borthwick, K. (2021). Researching massive open online courses for language teaching and learning. *ReCALL, 33*(2), 107–110. https://doi.org/10.1017/S0958344021000094

McCarthy, M., & O'Dell, F. (2005). *English collocations in use*. Cambridge University Press.

Moon, R. (2010). What can a corpus tell is about lexis? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 345–358). Routledge.

Nesselhauf, N. (2005). *Collocations in a learner corpus*. John Benjamins Publishing.

O'Keeffe, A. & McCarthy, M. (2010). Historical perspective: What are corpora and how have they evolved? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 3–13). Routledge.

Otto, P. (2021). Choosing specialized vocabulary to teach with data-driven learning: An example from civil engineering. *English for Specific Purposes, 61*, 32–46. https://doi.org/10.1016/j.esp.2020.08.003

Oxford Learner's Dictionary. (Online). https://www.oxfordlearnersdictionaries.com/

Palmer, F. R. (1997). *Semantics.* Cambridge University Press.

Partington, A. (1998). *Patterns and meanings. Using corpora of English language research and teaching.* John Benjamins.

Pellicer-Sánchez, A., Siyanova-Chanturia, A., & Parente, F. (2022). The effect of frequency of exposure on the processing and learning of collocations: A comparison of first and second language readers' eye movements. *Applied Psycholinguistics, 43*(3), 727–756. https://doi.org/10.1017/S014271642200011X

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), 10–24. https://so04.tci-thaijo.org/index.php/LEARN/article/view/111700

Phoocharoensil, S. (2020a). Collocational patterns of the near-synonyms *error*, *fault*, and *mistake*. *The International Journal of Communication and Linguistic Studies, 19*(1), 1–17. https://doi.org/10.18848/2327-7882/CGP/v19i01/1-17

Phoocharoensil, S. (2020b). A genre and collocational analysis of *consequence*, *result*, and *outcome*. *3L: The Southeast Asian Journal of English Language Studies, 26*(3), 1–16. http://ejournals.ukm.my/3l/issue/view/1326

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

Phoocharoensil, S. (2021b). Semantic prosody and collocation: A corpus study of the near-synonyms *persist* and *persevere*. *Eurasian Journal of Applied Linguistics, 7*(1), 240–258. https://doi.org/10.32601/ejal.911269

Phoocharoensil, S. (2023). Teaching English conditionals through data-driven learning (DDL): Perspectives of in-service EFL teachers. *Theory and Practice in Language Studies, 13*(10), 2620–2628. https://doi.org/10.17507/tpls.1310.21

Phoocharoensil, S. (2025). Synonymy in context: Analyzing the usage of *coming* and *upcoming* in American English. *The New English Teacher, 19*(2), 135–155. https://www.proquest.com/openview/574c9f1897d5495bd956bf0cf564e8fa/1.pdf?pq-origsite=gscholar&cbl=4531125

Piao, S., Bianchi, F., Dayrell, C., D'Egidio, A., & Rayson, P. (2015). Development of the multilingual semantic annotation system. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human*

Language Technologies (NAACL HLT 2015), Denver, Colorado, United States, pp. 1268–1274.

Roohani, A., & Heidari Vincheh, M. (2023). Effect of game-based, social media, and classroom-based instruction on the learning of phrasal verbs. *Computer Assisted Language Learning, 36*(3), 375–399. https://doi.org/10.1080/09588221.2021.1929325

Rungrueang, T., Boonprasert, P., Poempongsajaroen, S., & Laosrirattanachai, P. (2022). Corpus-based approach to generate a word list for food service. *THAITESOL Journal, 35*(1), 57–76. https://so05.tci-thaijo.org/index.php/thaitesoljournal/article/view/258591

Saeedakhtar, A., Bagerin, M., & Abdi, R. (2020). The effect of hands-on and hands-off data-driven learning on low-intermediate learners' verb-preposition collocations. *System, 91*, 102268. https://doi.org/10.1016/j.system.2020.102268

Schmitt, N. (2010). *Research Vocabulary: A Vocabulary Research Manual*. Palgrave Macmillan.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*(4), 484–503. https://doi.org/10.1017/S0261444812000018

Sridhanyarat, K., & Phoocharoensil, S. (2023). A corpus-based investigation of English near-synonyms: *Assess*, *evaluate*, and *measure*. *Humanities, Arts and Social Sciences Studies, 23*(1), 208–219. https://doi.org/10.14456/hasss.2023.19

Sumonsriworakun, P. (2022). A corpus-based investigation of English synonyms: *Disadvantage*, *downside*, and *drawback*. *LEARN Journal: Language Education and Acquisition Research Network, 15*(2), 649-678. https://so04.tci-thaijo.org/index.php/LEARN/article/view/259944

Szudarski, P. (2018). *Corpus linguistics for vocabulary: A guide for research.* Routledge.

Taylor, J. R. (2002). Near synonyms as co-extensive categories: 'high' and 'tall' revisited. *Language Sciences*, *25*(3), 263–284. https://doi.org/10.1016/S0388-0001(02)00018-9

Timmis, I. (2015). *Corpus linguistics for ELT: Research and practice.* Routledge.

Webb, S., Newton, J., & Chang, A. (2012). Incidental learning of collocation. *Language Learning, 63*(1), 91–120. https://doi.org/10.1111/j.1467-9922.2012.00729.x

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

Wynne, M. (2005). Stylistics: Corpus approaches. In E. K. Brown & A. Anderson (Eds.), *Encyclopedia of language & linguistics* (pp. 223 – 226). Elsevier.

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

Yang, C. T., Chen, H. H., Liu, C. Y., & Liu, Y. C. (2020). A semi-automatic error retrieval method for uncovering collocation errors from a large learner corpus. *English Teaching & Learning*, *44*(1), 1–19. https://doi.org/10.1007/s42321-019-00037-y

Yeh, E., & Mitric, S. (2023). Social media and learners-as-ethnographers approach: increasing target-language participation through community engagement. *Computer Assisted Language Learning, 36*(8), 1558–1586. https://doi.org/10.1080/09588221.2021.2005630

Yoo, I. W., & Shin, Y. K. (2020). Determiner use in English quantificational expressions: A corpus-based study. *TESOL Quarterly, 54*(1), 90–117. https://doi.org/10.1002/tesq.539