Alanazi *Asian. J. Second. Foreign. Lang. Educ.* (2022) 7:15
https://doi.org/10.1186/s40862-022-00138-5

Asian-Pacific Journal of Second and Foreign Language Education

RESEARCH Open Access

# Corpus-based analysis of near-synonymous verbs

Zaha Alanazi*

*Correspondence:
z.alonazi@mu.edu.sa
English Department, Majmaah University, Al Majma'ah, Saudi Arabia

**Abstract**

Despite having different semantic profiles, near synonyms are usually presented in dictionaries as being contextually interchangeable, which may lead EFL learners to assume their contextual interchangeability. Nevertheless, there is a scarcity of studies on how near synonyms are similar or different in their semantic and grammatical preferences. To enrich the literature on near synonyms' semantic and grammatical profiles, this study explores the collocational behaviors and the semantic preferences of the near-synonymous verbs (*affect* vs. *impact*). Sketch Engine was used to examine lexical collocates, the colligational profile and the semantic prosody of the two verbs. The findings revealed fine-grained contextual differences in their collocational, grammatical, and semantic preferences. Applications of the findings for English language teaching will be discussed a long with recommendations for future research.

**Keywords:** Sketch engine, Collocation, Synonyms, Colligation, Semantic, Corpus

## Introduction

This paper investigates the collocational behavior and semantic preferences of near-synonymous words. Despite its ubiquity in language, the linguistic phenomenon of synonymy is relatively under-researched compared to investigations of general lexical items (Edmonds & Hirst, 2002; Xiao & McEnery, 2006). Unlike absolute synonyms, which are rare in language, near- synonyms are widely common (Inkpen & Hirst, 2006; Murphy, 2003). Although collocations of lexical items have received considerable attention in empirical research recently, less research has examined the collocational and colligational patterns of near-synonymous words. Near-synonymous words, particularly verbs, are problematic for L2 learners not only because of their ubiquity, but also because they have similar connotational meaning, though they are not collocationally interchangeable (Liu, 2010; Yang et al., 2020). Subsequently, it is not unusual for ESL teachers to find student errors that can be attributed to the erroneous substitutions of near-synonymous lexical items (Chan, 2010; Nguyen & Webb, 2017). The knowledge of the subtle difference among near synonyms becomes more pressing if one accepts that knowing a lexical item requires knowledge of its collocates, colligates, and semantic preferences (Sinclair, 1998; Stubbs, 2013).

Sinclair (1991, 1998) argued forcefully that the ambiguity of a text is a result of a faulty focus on individual lexical items. In natural settings, people process language holistically, utilizing lexical and grammatical cues. As such, knowledge of collocations reflects one's fluency in using an L2. However, what actually constitutes collocates is still a matter of controversy in second language acquisition research. Sinclair (1991) describes two principles responsible for language organization at the phrase level: open choice-principle and idiom, or collocational principle. The first perceives language as "a series of slots which have to be filled from the lexicon" (Sinclair, 1991, p. 109) where the only choice constraint is grammar. However, it needs no proof that grammar is not the only controlling factor to our lexical choices, as language users may opt to use prefabricated structures "that constitute single choices, even though they appear to be analyzable into segments" (Sinclair, 1991, p. 110). The tendency to use prefabricated structures is understandable in light of the fact that despite the vastness of human memory capacity, the speed of processing capacity is limited, and hence, there is a need for prefabricated structures to shorten processing time (Nattinger & DeCarrico, 1992). One can imagine, however, the challenge of L2 learners acquiring collocational knowledge if we acknowledge that there is no total agreement even among native speakers on acceptable collocations (Partington, 1998) and that words are not only primed by other lexical items, but also grammatical categories and semantic preferences. Concordance-based studies have shown that a word may be primed by particular grammatical categories i.e., colligation (Nation, 2013) and that words with a wide range of collocates exhibit semantic preferences. Sinclair (2004) defines semantic preference as the tendency of lexical items to predominantly co-occur with lexical items of a particular semantic field. For example, the word preserve found to collocate mainly with abstract notions signifying importance such as integrity, unity, and anonymity (Li, 2019). Concordance data have also shown that in their semantic preferences, words often show favorable (positive) or unfavorable (negative) connotations (e.g., *outbreak* tends to collocate with negative notions like *conflict, disease,* or *violence*) (Partington et al., 2013, p. 81). Similarly, *cause* tends to collocate with unpleasant notions like *accident,damage, concern, disease,* and *death* (Stubbs, 1995). The tendency of words to appear in positive or negative contexts is referred to as semantic prosody (Stewart, 2010). Stewart contended that all words with the exception of grammatical words have the potential of having semantic prosody.

The collocational knowledge of lexical items can then be overwhelming to L2 learners, and near-synonymous lexical items can be even more confusing. Laufer (1990) pinpointed that one reason why synonyms contribute to the difficulty of acquiring vocabulary is that learners often substitute words with their synonyms without considering their collocational patterns. Awareness of lexical items semantic prosody adds further burden to learners' efforts of using near synonymous words appropriately. In a study of 123 ESL learners with proficiency levels ranging from intermediate to high level, Dushku and Paek (2021) found that learners exhibited noticeable difficulty in producing appropriate semantic prosody compared to their ability of recognizing it.

Near synonyms, nevertheless, are often presented in dictionaries and thesauruses in a way that implies their interchangeability, and while lexical collocations of lexical items might be provided, the semantic preferences and the preferred syntactic structures are usually underrepresented or vaguely implied. Hence, one way to facilitate learning of

semantically similar words is to highlight the similarities and differences of their collocational behaviors. Accordingly, there is a need for more detailed research on how presumably synonymous words behave differently or otherwise similarly.

Therefore, this study aims to examine the collocates, grammatical patterns, and semantic preferences of two near -synonymous verbs: *affect* versus *impact*. The rationale of this selection is twofold: First, verbs, in general, and in near-synonymous cases, in particular, are a major source of errors in a second language (L2). Nesselhauf (2003) emphasized that the focus in L2 learning should be on verbs, as they constitute the major difficulty when using verb noun collocations. Partington (1998, p. 77) argued that "one promising area for analyzing semantic prosody is verb phrase collocations with favorable or unfavorable objects." Second, these pairs of near synonyms are defectively presented in dictionaries (in the online Collin dictionary, in particular). The grammatical patterns and semantic preferences of these verbs are vaguely introduced. Bearing this in mind, the first part of this paper will investigate what the literature has revealed about knowledge of lexical items. The second section will be devoted to discussion of definition and the criteria of selecting and the role of collocations in second language learning. This will be followed by the methodology and the major results of the analysis.

## Literature review

### Synonyms and near synonyms

Being a significant rhetorical tool, synonymy is a ubiquitous phenomenon in language (Edmonds & Hirst, 2002; Partington, 1998). Divjak (2006, p. 21) referred to synonymy as a phenomenon that describes "one and the same situation, they name it in different ways, they represent it from different perspectives." This perspective supports the rarity of absolute synonyms because, as Cruse (1986, p. 270) puts it, "natural languages abhor absolute synonyms just as nature abhors a vacuum." On the other hand, near synonyms (i.e., words that share one or more semantic sense[s]) are widely used. Despite their pervasiveness, near synonyms are confusing, especially to L2 learners, as they are not contextually interchangeable and, hence, substituting one word with another may lead to unintended implications (Edmonds & Hirst, 2002). Dismantling fine-grained differences or similarities of near synonyms requires an in-depth examination of their use in various contexts given that knowing the meaning of a lexical item, as will be discussed below, goes beyond knowing its core semantic meaning.

### The meaning of lexical items

Sinclair (1998) among other researchers (Harmon et al., 2000; Hoey, 2005; Nation, 2013) criticized the traditional ways of presenting lexical words' meanings. He argued that the meaning of a word goes beyond its core meaning (paradigmatic level), calling for a top-down presentation of meaning (syntagmatic level), and contending that the paradigmatic view of meaning gives a word too much independency, which can lead to erroneous observation of meaning. It is this limited view of meaning rather than the textual structure that contributes most to the perceived ambiguity of a text. In a syntagmatic view, by comparison, the meaning is restrained by contextual factors. As such, a comprehensive view of meaning obliges consideration of both syntagmatic and paradigmatic views. The meaning of a lexical item, according to Sinclair, is composed of five

components including two obligatory and three optional yet genuine categories. The obligatory involves the core meaning and semantic prosody of a word. The optional categories involve collocation, colligation, and semantic preferences, which Sinclair (1998, p. 14) describes as "coordinated secondary choices within the item, fine tuning the meaning and giving semantic cohesion to the text as a whole". These three components were also emphasized by Hoey (2005, p. 116) in his lexical priming theory; "I would hypothesize that all words are primed for one or more collocations, semantic associations, and colligations, even if these are on the face of it unremarkable." With the absence of a universal definition, one needs to review the various definitions and classifications of collocations in second language acquisition (SLA).

#### Collocations and colligations

Collocation refers to lexical co-occurrence, and one of the earliest definitions of collocations was put forth by Firth (1957, p.11), stating "you shall know a word by the company it keeps.". Collocations, thus, range from strict or fixed combinations, such as *nice to see you* (Wray, 2002), to less fixed phrases (e.g., *completely different/new/free*) (Granger, 1998, p. 146). They also differ in their size (the number of words in a sequence), type (e.g., content words' collocation with function words, like *look at*, and content words collocating with content words, like *commit a suicide*), and in the range of potential collocates, as some words may have broader collocates than others (Nation, 2013). overall, the definitions of collocations can be grouped based on the selection based criteria into: statistical, semantic compositionality or colligational based definitions.

#### Statistical measures

Statistically- based category emphasizes form-focused or syntagmatic relations in a text i.e., words are considered collocates if two or three words co-occur within a particular span from each other (Sinclair, 1991; Stubbs, 1995). Within in this category, two main approaches by corpus-based language studies are used: frequency based and strength of association-based measures. Absolute (raw) frequency which relies on counting the instances of co-occurrence of word combinations has shown to be an informative measure for studies on relationship between the frequency of word combinations and the psychological processes of language learning such as retrieving, noticing…etc. (Ellis, 2002). Nation (2013) stressed the pedagogical benefit of word frequency, as learners need to learn items they will encounter and use most often. Although absolute frequency can play a part in psychological processes of language learning, it may not reflect the regularities of usage of a language.

Studies have shown that collocates with high frequency scores in specific corpus may not necessarily be frequent or regular in a language as frequency scores might be inflated by the overuse in specific texts or a small number of speakers/writers. For instance, Gablasova et al. (2017) found that risk issues and moral issues have comparable raw frequency, with 54 and 51 occurrences, respectively. However, all the 54 instances of the first expression were from one text while the latter occurred in over 41 texts.

The second statistical approach to identifying collocates focuses on the strength of the association between words. Three main indices have been used in corpus- based research: MI scores, T-score and LogDice. MI or the mutual information index (MI)

which has been used in numerous research studies (Hunston, 2002; Bestgen & Granger, 2014), is used to mathematically express the ratio between the frequency of the collocation and the frequency of random co-occurrence of the two words in the combination (Church & Hanks, 1990). While MI score can be used to indicate how strongly words are associated, it can also yield items that are strongly correlated, but hardly used in language (Bestgen & Granger, 2014; Nation, 2013).

The second commonly used score for collocation is the T-score which is also designated as a measure of "certainty of collocation" (Hunston, 2002, p. 73). T-score found to yield similar results to raw frequency measures (Durrant & Schmitt, 2009). Nevertheless, while collocates with high t-scores are frequent in language, not all frequent collocates have high t-scores. Another downside of t-score is its bias to corpus size making their scores inappropriate for comparison of collocates across corpora of various size (Gablasova et al., 2017).

LogDice is another measure of strength and although similar to MI, LogDice has not been explored by language learning research (Gablasova et al., 2017). In their critical review of measures of the strength of collocations, Gablasova et al. (2017) concluded that LogDice is preferable to MI as it provides standardized measure with a maximum value of 14, making it comparable across corpora of different size. In addition, LogDice is a preferred measure with large corpora as scores of other traditional measures can be skewed when used with enormous size corpora.

In addition to statistics, Granger and Meunier (2008) called for utilizing corpora and teachers' senses in the selection of which collocates to teach. It is only through human intervention that one can decide the pedagogical value of collocations (Ackermann & Chen, 2013), since even what seems to be statically significant collocations may not be frequently used in multiple contexts.

#### Semantic-transparency

The second criterion for defining collocations relies heavily on the semantic transparency of the collocates (i.e., the compositionality of the combinations compared to idioms). In this sense, collocations can be divided based on their compositionality into: (1) idioms where the meaning of combination cannot be deduced from the constituents, (2) figurative where the meaning is figuratively expressed, and (3) literals where the meaning of the parts is transparent enough to contribute to the meaning of the whole (Nation, 2013).

#### Colligational Profile

The third definition is based on the syntactic behavior of a word. Hoey (2005) contended that just as a word is primed to collocate with another lexical item, it also tends to avoid co-occurrence with a particular grammatical function. Unlike collocations, colligation refers to the "co-occurrence of a member of a grammatical class—say a word class—with a word or phrase" (Sinclair, 1998, p. 15). Hoey refers to the tendency of grammatical co-occurrence as positive colligations whereas negative colligations refer to the avoidance of particular grammatical structures (e.g., *consequence* is found to have negative colligation with the object function). Drawing on this premise, Hoey proposed three components for colligations:

1. The grammatical company a word or a word sequence keeps, either within its own group or at a higher rank.
2. The grammatical functions preferred or avoided by the group in which the word or word sequence participates;
3. The place in a sequence that a word or a word sequence prefers (p. 43).

Hoey (2005) also emphasized domain-based priming, or the notion that what a word primes in one context may be different in another. The different approaches to collocations imply an increasing interest in their significance for language learners, as will be illustrated in the coming discussion.

#### Pedagogical importance of collocations

Research in first language (L1) and L2 learning has revealed the prevalence of phraseology in both spoken and written forms. Multi-unit words help learners enhance their listening and reading comprehension, along with accuracy and fluency, in both written and oral production (Granger & Meunier, 2008). In an in-depth review of studies of formulaic sequence in L1s and L2s, Conklin and Schmitt (2012) found compelling evidence that native speakers tend to process, access, and produce formulaic sequences faster than novel utterances. The criticality of frequently co-occurred words or collocation is manifested in the fact that correct use mirrors native-like competence and does correlate with human judgment of writing quality(Paquot, 2018) whereas erroneous uses of collocates "immediately unmask the non-native speaker" (Hindl, 2010, p. 47). Unlike idioms, collocations have a wider scope and are more prevalent in language, yet are illusive. Nation (2001, p. 324) argued that collocations can be unpredictable, both lexically and grammatically. This unpredictability is what makes collocations a contributor to advanced L2 students' errors (Osborne, 2008). Learners tend to treat multi-unit words as two separate constituents, leading to errors in grammatical feature transfer as in **natives speakers*. They also tend to associate lexical and grammatical components inappropriately (e.g., using *since* with *have*, even if the former instances does not have a temporal function (Osborne, 2008). Laufer and Waldman (2011) found evidence of underuse and erroneous use of verb-noun collocations across all proficiency levels of L2 learners compared to their native peers. In addition to learners' lack of collocational knowledge, part of the confusion in using near synonyms can be attributed to L1 transfer, particularly when L2 near synonyms have the same or similar equivalents in learenrs'L1 language (Chan, 2010; Liu & Zhong, 2016).

Multi-unit words are mainly problematic to learners whose primary input is text rather than speech, given that the text does not indicate that the individual components should be perceived as one chunk. Learners may notice unknown individual words, but hardly notice unknown chunks of language (Wible, 2010). Even if correctly used, learners tend to rely on small sets of collocations (Granger, 1998).

Nation (2013) stressed that the combination of collocations is not random, because collocates are fulfilling semantic and grammatical functions. Thus, they are problematic from a decoding perspective, because the meaning of a word is determined by the company it keeps. Collocations also cause difficulty from a typological perspective, because collocations are not as homogeneous as they are perceived but rather encompass a

variety of word combinations. In alliance with Nation's (2001) contention, Howarth (1998) found that most L2 learners' collocation errors were from collocations that allow some substitution of both elements. He attributed non-native speakers' (NNSs') errors in using collocations to teaching strategies that focus on the grammaticality of words' combination and to teachers' lack of knowledge of "phraseological mechanisms of language" (p. 186). Revealing useful information about the differences between words, Nation (2013) emphasized the significance of investigating semantic preferences and grammatical patterns of collocations.

## Motivation and study questions

Despite the importance of collocations, it was not until recently that researchers turned their attention to developing lists of corpus-based collocations rather than lists of individual words (Ackermann & Chen, 2013; Durrant, 2009; Ellis et al., 2008; Simpson-Vlach & Ellis, 2010). Durrant's (2009) corpus-based list is mainly of grammatical collocation, or of closed-class collocations (e.g., determiners or prepositions plus a noun). Ellis et al. (2008) developed the academic list of formulas of academic and spoken language. N-grams were extracted from spoken corpora in MICASE and BNC corpora. Written corpora were collected from Hyland's research article corpora and some academic articles from the BNC. Simpson-Vlach and Ellis (2010) used both statistics and human judgment to identify collocates. They used a weighted combination of MI and frequency; their list, however, includes items that are not "grammatically well-structured" (Nation, 2013, p. 496). Ackermann and Chen (2013) developed a list of cross-disciplinary lexical collocations. The two-word collocations were extracted from the written curricular component of the 25 million word Pearson Corpus of Academic English (PiCAE) and were restricted to open class collocates (including no functional words). Similar to Simpson-Vlach and Ellis, the selection of collocations was based on statistical frequency, though, human judgment was used for both the selection and final refinement of collocational lists.

A critical view of the aforementioned studies revealed that the there is a tendency to compile lists of collocations or phrases with little consideration, if any, to the problematic words that are particularly synonymous or near-synonymous verbs. Words of similar meaning are more difficult to learn than words that are not semantically related (Waring, 1997). Moreover, semantic preferences, colligational patterns, and domain-specific collocations were largely overlooked. These limitations call for more detailed quantitative and quantitative analyses of the problematic verbs that have similar senses. Therefore, this study attempts to enrich the research on collocations by shedding light on the collocational behaviors of the near-synonymous verbs *affect* versus *impact*. Stewart (2010) emphasized that corpus-based studies can be sometimes provoked by intuition and introspection. Hence, the selection of verbs in this study is triggered by my intuition, as a non-native speaker of English and as an ESL teacher, of the type of verbs that might be problematic to master, especially near-synonymous words. In the entry of the verb *impact*, the Merriam Webster dictionary lists *affect* as the first synonym. Conversely, the first entry meaning of *affect* defines it as "to act upon (a person or a person's feelings) so as to cause a response" and presents *impact* as the first synonym. Collins Online Dictionary (American English) defines *affect* as" If something affects a

person or thing, it influences them or causes them to change in some way". It also shows *impact* to be among the first listed synonyms of *affect.* In the usage notes of *impact* as a verb in Meriam Webster dictionary, the following example was given "This need to hold stock for 12 months will *impact* mutual funds". It needless to say, that the verb *affect* can meaningfully replace *impact* in this context "…affect mutual funds". Nevertheless, this does not mean that that the two verbs can be used interchangeably in all contexts, nor does it exclude the fact that one verb is more preferably used in a context than the other.

The dictionary definition then, introduces the two verbs as if they are contextually interchangeable. It is only through a detailed description of each pairs' collocational behavior, one can pinpoint similarities and differences. Another major motive for selecting the current set of near-synonymous verbs is that they both share the same equivalent in Arabic (the author's native language) making them more susceptible to erroneous use by EFL Arab learners. Hence, following Hoey's emphasis on the three main elements of colligational analysis and Nation's (2013) recommendations, this study will examined both the lexical collocations in terms of the type of lexical collocates that a verb primes, the grammatical categories or functions that the collocate words tends to take and the potential semantic preferences of the near-synonymous verbs. Since the two verbs have more than one meaning in the dictionary and to arrive at a meaningful comparison, the analysis will focus on collocates with which the near synonymous are used in their shared sense i.e., producing an effect upon (someone or something). In particular, this study attempts to answer the following question:

*Q1. Are their differences between the adverbial lexical collocations of the near-synonymous verbs affect versus impact in terms of their frequency, semantic meaning, connotative meaning and their preferred syntactic positions (e.g. post or premodifiers) ?*
*Q2. Are their differences between nominal lexical collocations of the near-synonymous verbs affect versus impact in terms of their frequency, their semantic meaning, their connotative meaning, and their preferred syntactic positions?*

#### Analysis tools and corpus

The tool Sketch Engine was used as it has several integrated functions for linguistic analysis. The tool includes various corpora that collectively contain more than 500 million words from different languages. The corpus used for the current analysis is the written texts in the British National Corpus (BNC). BNC contains more than 100 million words from texts from different genres (newspapers, textbooks, other media) representing British spoken and written English from the latter part of the twentieth century.

Sketch engine contains several functions that can be used for textual analysis, one of which is "Word Sketch" which was used to answer the research questions of the current study. The Word Sketch (WS) tool provides a summary of the strongest collocates of a word or a phrase and displays them as sorted by grammatical relation. For example, the output could be sorted by words that usually appear as modifiers objects/subjects of the targeted verb.

The function generates a summary list of the collocates of near-synonymous words arranged by their grammatical categories (e.g., collocates in the subject or the object

position for both near-synonymous words). This function offers a dropdown list for specifying the part of speech and another for specifying the sub-corpora. Collocates in each column are presented with their frequency scores and are sorted in descending order based on their typicality score (another term for LogDice score used by Sketch Engine). The terms LogDice and typicality score are used interchangeably in this paper.

#### Analysis of data

Both quantitative and qualitative analysis were used simultaneously in this study. The first phase compared the frequency of the two near-synonymous verbs in the BNC written texts. Concordance lines were examined qualitatively for each verb to exclude frequency counting of irrelevant examples (i.e., the verbs *affect/impact* being used with a meaning different than the one being investigated). The second step focused on examining the generated list of collocates to select the top 30 collocates that met the inclusion criteria: (a) the nominal and adverbial lexical collocates of the near-synonymous verbs that are content words, (b) collocate content words co-occur with *impact/affect* as near-synonymous verbs (i.e., the verb means producing an effect upon someone or something). The exclusion of function words was driven by the fact that functional words add little to the semantic comparison of the collocates of the two near-synonymous verbs, which was the focus of this study. Excluding content collocates with irrelevant sense meant arriving at more reliable and reasonable findings since words based on their semantic senses may have different collocational profiles.

Based on the exclusion criteria, function words that include auxiliary verbs, prepositions, articles, conjunctions, linking adverbials, adverbials of frequency (e.g., *always, sometimes*) and pronouns were not included. Also, the concordance lines of the content collocate were examined to ensure that that the generated collocates are collocates of the verb with the intended meaning. In fact, this was not an issue with *affect* collocates, as all examples reflected the intended meaning. However, as will be shown later in the analysis, the generated list of *impact* collocates included numerous examples of use irrelevant to the shared sense being examined.

The analysis indicated earlier focused on nominal and adverbial collocates. Nominal collocates include those in the subject, object categories. Adverbial collocates refer to adverbials modifying the verb. Using the WS tool, I first searched for *affect* and specified the part of speech as a verb and in the inquiry box of sub-corpus and selected written texts from the dropdown list. The auto frequency and a minimum typicality score of 0 (LogDice score), which was part of the default setting were kept unchanged. To facilitate qualitative analysis of the collocates and to group them into lexical groups, only the top 30 with the highest LogDice score (the default measure used in Sketch Engine) in nominal and adverbial categories were considered; if items were excluded for not meeting the inclusion criteria (e.g., function words), the analysis extended beyond the 30 collocates (in case the list had more than 30 items) to compensate for the excluded ones until the list had no less than 30 lexical content words.

After refining the list of top 30 collocates, they were categorized based on their semantic meaning into semantic sets or thematic groups. Semantic sets refer to "items which share a semantic feature, for example that they are all about, say, sport or suffering" (Sinclair, 2004, p. 142).

## Results and discussion

### Frequency and adverbial collocates

#### *Affect/impact*

The examination of frequency for *affect* revealed that *affect* was way more frequently used in the BNC written texts, with 12,324 occurrences (122.57 per million tokens) compared to 152 reduced to 107 instances of impact (1.51 per million tokens) after eliminating the examples with irrelevant meaning to the one under investigation. Table 1 below illustrates some of the excluded instances of *impact*. It can be noticed that the verb is used to mean hitting forcefully or to refer to senses that are not relevant to the current analysis.

To examine potential differences in nominal and adverbial lexical collocates of *affect* and *impact*, the function WS was used. Table 2 shows only the top adverbial collocates from the first 30 collocates generated by the program (the list of all adverbial collocates is shown in "Appendix A"). Three adverbs (*this, so, similarly*), which did not meet the criteria, were excluded.

The generated list of adverbial collocates shows a variation in their frequency, ranging from 8 to 307 instances reported for *adversely,* placing it as the most frequent adverbial collocate for *affect*. Differences in the strength of association were also noticed with *adversely* having the highest LogDice score of 11.3 out of 14., the maximum score possible, while the adverb *especially* reported the lowest with 5.94.

A qualitative analysis was conducted on the adverbial collocates of *affect* and three semantic or thematical categories were identified, as shown in Table 3: (1) adverbs denoting intensity, (2) adverbs denoting type/specificity, and (3) adverbs denoting possibility. Each collocate was then assigned to its relevant category.

The semantic grouping of adverbial collocates of *affect* shows that the majority of its adverbial modifiers describe a degree of intensity. These adverbials can be further classified based on their connotative meaning into modifiers with strong negative connotations (*adversely, badly, directly, seriously, worst, severely*), modifiers signaling power or significance (*profoundly, greatly, deeply, radically, disproportionality, strongly, mainly, considerably, powerfully*) and modifiers indicating lack of strength (*marginally, partially, little*).

**Table 1** A sample of excluded instances of the verb impact

| | | |
|---|---|---|
| Paper until his embarrassment had faded. Then he began to tap: '1. The range from the firing point to where the rounds | Impacted | Was between 35 m (the shot that hit Barling) and 30 m (the ones that hit the pillar)…. Twenty minutes later |
| Below her was the corpse of a woman. Tallis had seen the grimacing features as she was carried to the grave. Now, as she | Impacted | With the body, she felt the bones stir. A sap rose in her, human warmth in the veins of the wood. The dull, meaningless |
| The molecules and ions in gases and liquids are in a state of constant motion. By | Impacting | With neighbouring particles they vibrate about a locus, and only appear to remain in a fixed position. This movement on |
| And the like. They avoid places where they perceive the risk of assault to be high. They are extremely vulnerable to | Impact | By vehicles, although they will trade off this risk against increased journey length: it is not unusual to see people |
| Where is she?<br>MAX: She should be at the dentist's all day tomorrow. ABBERLEY: Her teeth are perfect.<br>MAX: She has four | Impacted | Wisdom teeth. ABBERLEY: But no decay? MAX: I'll ask her, if you like |

**Table 2** Top 30 adverbial collocates of the verb affect

| | | | |
|---|---|---|---|
| 1 | Adversely | 307 | 11.26 |
| 2 | *Directly* | 146 | 8.83 |
| 3 | Badly | 110 | 8.81 |
| 4 | Seriously | 106 | 8.73 |
| 5 | Significantly | 84 | 8.56 |
| 6 | Severely | 62 | 8.46 |
| 7 | Profoundly | 45 | 8.44 |
| 8 | *Materially* | 25 | 7.76 |
| 9 | Greatly | 42 | 7.47 |
| 10 | *Indirectly* | 22 | 7.46 |
| 11 | Deeply | 38 | 7.38 |
| 12 | *Particularly* | 63 | 7.06 |
| 13 | Radically | 16 | 6.91 |
| 14 | Inevitably | 20 | 6.8 |
| 15 | Disproportionately | 13 | 6.8 |
| 16 | Strongly | 26 | 6.65 |
| 17 | Worst | 10 | 6.49 |
| 18 | Little | 34 | 6.32 |
| 19 | Mainly | 21 | 6.31 |
| 20 | Considerably | 14 | 6.23 |
| 21 | Powerfully | 9 | 6.21 |
| 22 | Substantially | 12 | 6.2 |
| 23 | Drastically | 9 | 6.19 |
| 24 | Vitally | 8 | 6.1 |
| 25 | Marginally | 8 | 6.07 |
| 26 | Immediately | 22 | 6.04 |
| 27 | Critically | 8 | 6 |
| 28 | Potentially | 11 | 6 |
| 29 | Markedly | 8 | 5.97 |
| 30 | Especially | 14 | 5.94 |

**Table 3** Lexical grouping of the adverbial collocates of the verb affect

| Lexical grouping | Number of collocates | Examples |
|---|---|---|
| Intensity/degree and emphasis | 26 | Adversely, directly, badly, seriously, significantly, severely, profoundly, greatly, indirectly, deeply, radically, inevitably, disproportionality, strongly, worst, immediately, little, mainly, considerably, powerfully, substantially, drastically, vitally, critically, marginally, markedly |
| Type/specificity | 3 | Materially, particularly, especially |
| Possibility | 1 | Potentially |

To a lesser degree, the verb *affect* collocates with adverbs indicating possibility (e.g., *potentially*) and with adverbs signaling type/specificity (*materially, particularly, especially*), respectively.

As for the syntactic placement of the adverbial collocates, the analysis of concordance lines showed that adverbial collocates of *affect* have a greater tendency to occur as premodifiers with over 95% of the top 30 adverbs occurring before *affect*. For

**Table 4** Examples of concordance lines of adversely

| Ecology of the waters have been *adversely* | Affected | By overfishing and the seals will starve |
|---|---|---|
| The properties themselves and *adversely* | Affect | The ability of individuals to sell them |
| Problems. Importantly they also *adversely* | Affect | Levels of amenity in both residential and |
| Company if those issues might *adversely* | Affect | A possible management buy-out; and they |
| Much in the early days, they may *adversely* | Affect | Their interests in the medium term |
| Any doubt given to the party *adversely* | Affected | See Chitty on Contracts, Chapter 14 |
| Shopping centre,for instance, will *adversely* | Affect | The tenant's business. 1.4"Works |
| Extension or reduction does not *adversely* | Affect | The tenant's use or occupation of the premises |

**Table 5** Examples of concordance lines of impact adverbial collocates

| | | |
|---|---|---|
| Microsoft's dominant strength rests with the desktop. Brown estimates that a loss on the server side would adversely | Impact | NT's position on the desktop. </s><s>'At the very least,' it says, 'Microsoft must enhance its credibility on the server side |
| The process in both branches of the profession | Impacts | Adversely on women and ethnic minorities, who, even when they do enter the profession, tend to be relegated to lower |
| Different kind from their middle class counterparts. Urban decay brings with it a host of associated miseries which | Impact | Especially severely on old people. We are now all familiar with the sadness and anxieties of old people left in derelict |
| Knowledge of the research practices and needs of a particular group, e.g. historians or environmentalists, can | Impact | Favourably on research in that area. An active policy of data identification and acquisition in areas should be |
| Were devalued. As a result of the Group's hedging policy, the benefits of these currency movements will partially | Impact | 1993 and will clearly be seen in 1994. Shareholders' funds have reduced by IR£37.4 million due to the loss for the period |
| Factors. However, it is important not to neglect the style and form of the implementation process itself which will | Impact | Directly upon those who receive services and will also determine how scarce resources are allocated. The |
| Show managers how changes in resources and priorities can lead to changes in output… Performance measures still | Impact | Only slightly on resource allocation decisions'. Certainly the measurement of programme expenditure has lagged far |
| The price indications in (i) above are based. </s><s>Clear indication should be given of any areas of uncertainty that may | Impact | Significantly on the amount of the offer; plans for [name] and its employees. The question of employee |
| By Friday 1st December, 19XX. Any conditions attaching to your indicative offer. Any areas of uncertainty that may | Impact | Significantly on the quantum of your offer. The extent of any due diligence procedures that you would wish to carry out |

example, only three instances of *adversely* out of 307 (see Table 4) and six instances of *directly* out of 146 instances were post-modifiers whereas all examples of *badly* were premodifiers (Table 5).

Contrary to *affect*, the verb *impact* was shown to have a smaller set of adverbial collocates. The generated list included 22 adverbial collocates total, six of which were excluded, two were linking adverbials, two adverb of frequency and two adverbs (*deeply, overhead*) were irrelevant examples in which *impact* was used in a sense not within the scope of the current analysis. Besides the small number of adverbial collocates (see Table 6), the scores of the strength of association were relatively lower than that of *affect* collocates. Except for the four top collocates whose typicality scores ranged from 8. to 6., the majority of *impact* adverbial collocates have typicality scores

**Table 6** Adverbial collocates of impact (verb)

| | | | |
|---|---|---|---|
| 1 | Differentially | 1 | 8.09 |
| 2 | Summarily | 1 | 7.46 |
| 3 | Adversely | 2 | 6.84 |
| 4 | Favourably | 1 | 6.07 |
| 5 | Partially | 2 | 5.46 |
| 6 | Progressively | 1 | 5.4 |
| 7 | Inevitably | 3 | 5.39 |
| 8 | Significantly | 4 | 5.13 |
| 9 | Severely | 2 | 4.87 |
| 10 | Dramatically | 1 | 4.52 |
| 11 | Strongly | 2 | 3.72 |
| 12 | Primarily | 1 | 3.55 |
| 13 | Directly | 2 | 3.23 |
| 14 | Greatly | 1 | 2.95 |
| 15 | Slightly | 1 | 2.17 |
| 16 | Certainly | 1 | 1.48 |

between 5.4 and 1 (i.e., lower than the minimum LogDice score of affect adverbial collocates, which was almost 6.).

The lexical grouping of impact adverbial collocates resulted in two semantic groupings: intensity/emphasis and gradation. Almost all adverbial collocates, except *progressively,* belong to the intensity category, hence, overlapping with that of *affect* adverbial collocates. Nevertheless, out of the shared adverbs (e.g. *adversely, partially, inevitably, significantly, dramatically, severely, strongly*) only two have a negative connotation (*adversely, severely*). These findings are supported by frequency counts and LogDice scores of adverbials from the intensity category, suggesting that *affect* tends to prime more intensifying modifiers, particularly those of negative connotation, than its near-synonymous verb *impact.*

As for the syntactic placement of impact adverbial modifiers, the concordance lines revealed that compared to the adverbial collocates of *affect*, the adverbial modifiers of *impact* tend to appear more frequently in a post-modifier position. In fact, only six out of the 16 adverbial collocates of *impact* appeared in a pre-modifying slot.

### Nominal subject collocates

#### *Affect*

The nominal subject collocates of *affect* were shown to have 4492 instances in the written corpus. The qualitative analysis of the first top 30 collocates revealed that all subject collocates were inanimate abstract nouns. Although the analysis focused on the top 30 collocates shown in Table 7, a quick look at the rest of the subject collocates point to a similar conclusion (see "Appendix B"). The analysis also revealed five semantic categories: cognition and aptitudes, action/behavior and motion, life and environment, law and order, problems/issues (see Table 8). The semantic group labeled problem/issues (e.g. HIV, pollution, recession, animosity) and the law order category provide further support for the findings from adverbial collocates that *affect* appears in more authoritarian and/or negative contexts.

**Table 7** Nominal subject collocates of affect

| Rank | Collocate | Raw frequency | LogDice score |
|---|---|---|---|
| 1 | Factor | 182 | 9.53 |
| 2 | Change | 151 | 8.87 |
| 3 | Decision | 86 | 8.21 |
| 4 | Issue | 76 | 8.11 |
| 5 | Matter | 60 | 7.98 |
| 6 | Recession | 38 | 7.79 |
| 7 | Disease | 41 | 7.72 |
| 8 | Condition | 47 | 7.54 |
| 9 | HIV | 20 | 7.12 |
| 10 | Legislation | 27 | 7.1 |
| 11 | War | 39 | 7.09 |
| 12 | Problem | 56 | 6.94 |
| 13 | Action | 32 | 6.93 |
| 14 | Crisis | 21 | 6.91 |
| 15 | Influence | 20 | 6.87 |
| 16 | Policy | 35 | 6.81 |
| 17 | Event | 28 | 6.78 |
| 18 | Aids | 16 | 6.76 |
| 19 | Law | 34 | 6.72 |
| 20 | Way | 26 | 6.68 |
| 21 | Trend | 16 | 6.57 |
| 22 | Animosity | 13 | 6.55 |
| 23 | Climate | 14 | 6.54 |
| 24 | Pollution | 14 | 6.52 |
| 25 | Development | 23 | 6.52 |
| 26 | Process | 26 | 6.51 |
| 27 | Consideration | 16 | 6.51 |
| 28 | Illness | 14 | 6.49 |
| 29 | Proposal | 18 | 6.42 |
| 30 | Presence | 14 | 6.4 |

**Table 8** Semantic categories of nominal subject collocates of affect

| Semantic category | Number of collocates | Examples |
|---|---|---|
| Cognition and attitudes | 9 | Factor, change, issue, decision, matter, influence, consideration, trend, way |
| Action/behavior/process | 8 | Action, development, process, events, aid, proposal, presence, activity |
| Life/environment | 1 | Climate |
| Problems/issues | 9 | War, HV, pollution, recession, disease, problem, crisis, animosity, illness |
| Law/order | 3 | Legislation, policy, law |

#### *Impact*

Compared to *affect*, the subject collocates of *impact* (see Table 9) are relatively small, with only14 instances reduced to nine after eliminating irrelevant collocates (*plane, plume, goblin, lightening, road*). Similar to its adverbial collocates, there is a relatively lower association between *impact* and its subject collocates compared to *affect*

**Table 9** Subject collocates of impact

| Rank | Collocate | Raw frequency | LogDice score |
|---|---|---|---|
| 1 | Initiative-which | 1 | 10.8 |
| 2 | Profession | 1 | 5.27 |
| 3 | Activity | 2 | 4.12 |
| 4 | Technology | 1 | 3.82 |
| 5 | Process | 2 | 3.76 |
| 6 | Section | 1 | 3.38 |
| 7 | Department | 1 | 2.94 |
| 8 | Fact | 1 | 2.68 |
| 9 | Price | 1 | 2.61 |

collocates. Except for the word *initiative*, all the generated subject collocates have a typicality score less than six, which is the lowest score of *affect* subject collocates, indicating that subject collocates of *impact* co-occur frequently with other words in the language. In addition, the small frequency counts of the collocates (1–2) indicate their scarcity in written discourse. As for shared collocates, the examination of the generated list revealed that, minus the exception of the word *process*, no shared subject collocates were found between *affect* and *impact*. Furthermore, similar to subject collocates of *affect,* all subject collocates of *impact* are abstract entities.

Despite having small number of collocates, the very low frequency and the miscellaneous semantic fields of *impact* subject collocates make it difficult to assign them into semantically parsimonious categories. Overall, four categories can be identified: cognition and aptitudes (*fact*), action/process (*initiative, activity, process*), business/technology (*price, technology*), and institutions/divisions (*profession, department/section*). The first two semantic categories overlap with that of *affect.* Nevertheless, the small number of collocates under cognition and action/process categories and no collocates related to problems/issues and law/order semantic groupings support findings from the analysis of adverbial collocates in that, unlike *affect*, *impact* is less likely to collocate with authority/law or problem-related words. While the shared semantic categories and the one shared collocate (*process*), suggest that both near-synonymous verbs can collocate with subject nominals relating to cognition and action, *impact* tends to demonstrate not only a diminished frequency, but also a weaker association with words related to cognition and action.

### Nominal object collocates

#### *Affect*

The quantitative analysis of object nominal collocates of *affect* revealed that nominal collocates are relatively more frequent in the object category with 8477 instances, almost double their subject counterparts (see the whole list in "Appendix B"). As shown in Table 10, the nominal collocates in the object category are mainly abstract nouns and, to a lesser extent, impersonalized animate nouns (people, *woman, individual, family and child*). The qualitive analysis of the top 30 collocates, as shown in Table 11, pointed to six thematical or lexical groupings: nature/environment, action/behavior, cognition and aptitudes, cause and relationship, degree/ quality, and trade/material.

**Table 10** Nominal object collocates of affect (Verb)

| Rank | Collocate | Raw frequency | LogDice score |
|---|---|---|---|
| 1 | Performance | 85 | 7.82 |
| 2 | Life | 145 | 7.78 |
| 3 | Health | 65 | 7.75 |
| 4 | Behavior | 73 | 7.67 |
| 5 | Area | 129 | 7.66 |
| 6 | Outcome | 60 | 7.63 |
| 7 | People | 171 | 7.12 |
| 8 | Attitude | 44 | 6.98 |
| 9 | Price | 60 | 6.98 |
| 10 | Level | 63 | 6.96 |
| 11 | Ability | 42 | 6.92 |
| 12 | Rate | 59 | 6.87 |
| 13 | Woman | 75 | 6.83 |
| 14 | Industry | 40 | 6.77 |
| 15 | Individual | 36 | 6.74 |
| 16 | Quality | 41 | 6.73 |
| 17 | Relationship | 46 | 6.7 |
| 18 | Trade | 32 | 6.65 |
| 19 | Business | 48 | 6.62 |
| 20 | Output | 30 | 6.6 |
| 21 | Economy | 32 | 6.58 |
| 22 | Decision | 51 | 6.55 |
| 23 | Country | 44 | 6.55 |
| 24 | Family | 39 | 6.51 |
| 25 | Environment | 31 | 6.51 |
| 26 | Result | 48 | 6.51 |
| 27 | Child | 69 | 6.49 |
| 28 | Right | 62 | 6.48 |
| 29 | Balance | 30 | 6.46 |
| 30 | Function | 34 | 6.45 |

**Table 11** Semantic categories of the object nominal collocates of affect

| Semantic category | Number of collocates | Examples |
|---|---|---|
| Cognition, attitudes, and aptitudes | 4 | Decision, attitude, function, ability |
| Action/behavior/motion | 2 | Behaviour, performance |
| Life/environment | 10 | Life, health, people, child, women, individual, area, country, family, environment |
| Cause and relationship | 4 | Relationships, output, outcome, result |
| Trade/business | 5 | Price, industry, business, trade, economy |
| Degree/quality | 5 | Level, rate, quality, balance, right |

#### *Impact*

As shown in Table 12, compared to *affect*, the verb *impact* has a smaller number of nominal object collocates (28 collocates) after excluding items not meeting the preset criteria. Similar to its adverbial and subject nominal collocates, almost half of

**Table 12** Object collocates of impact

| Rank | Object collocate | Raw frequency | LogDice score |
|---|---|---|---|
| 1 | ISV | 1 | 8.93 |
| 2 | CPU | 1 | 8.21 |
| 4 | Earning | 2 | 6.71 |
| 5 | Availability | 1 | 6.07 |
| 6 | Conduct | 1 | 5.9 |
| 8 | Margin | 1 | 5.64 |
| 9 | Perception | 1 | 5.36 |
| 10 | Screen | 1 | 5.23 |
| 11 | Performance | 3 | 4.66 |
| 12 | Vision | 1 | 4.55 |
| 13 | Means | 1 | 4.28 |
| 14 | Speed | 1 | 4.24 |
| 15 | Program | 1 | 4.17 |
| 16 | Employee | 1 | 4.1 |
| 17 | Offer | 1 | 3.9 |
| 18 | Contact | 1 | 3.4 |
| 19 | Research | 1 | 3.4 |
| 20 | Business | 2 | 3.1 |
| 21 | Father | 1 | 3 |
| 22 | Operation | 1 | 3 |
| 23 | Market | 1 | 2.7 |
| 24 | End | 1 | 2.5 |
| 25 | Change | 2 | 2.5 |
| 27 | Plan | 1 | 2.2 |
| 28 | Work | 1 | 0.8 |

*impact*'s object collocates show relatively lower association scores compared to their *affect* counterparts.

The nominal object collocates are mainly abstract, except for the words *employee* and *father*. They can mainly be categorized under the following lexical groups: action/process/behavior (*performance, availability, speed, operation, change, research*), cognition/aptitudes (*perception, conduct*), business/communication/technology (*isv, cpu, vision, screen, employee, program, earning, market, plan, offer,contact, means, margin,business, contact,end*), and people (*father*). It is worth mentioning that the collocates *isv, cpu,* and *vision* were categorized under business, because they were found to refer to commercial brands (see Examples 1–2 below). Similarly, the word *screen* was also used to refer figuratively to the business of broadcasting.

While categories of cognition, action, and people overlap with that of the *affect* object collocates, the near-synonymous verbs share only one object collocate: *performance.*

Comparing the findings from subject and object collocates of *impact* versus *affect* indicates that while the subject category for both verbs primes abstract nouns, the object category can include both abstract and concrete entities. Another noticeable difference between *impact* versus *affect* collocates is that while *affect* is used in a wider range of contexts, as reflected by the semantic grouping of its collocates,

*impact* seems to be used in relatively restricted contexts, mainly in business and technology. This is supported by the examination of concordance lines for *impact* nominal collocates, which were mainly business- or technology-related (see Examples 4–9).

### Examples

1. *Availability dates for OSF/1 on MIPS, he said, had been re-targeted until after the Alpha version, but as the developers version is already shipping, the date change should not impact ISV and customer development schedules too heavily.*
2. *For applications requiring thousands of input–output points, an intelligent controller and separate VME chassis, connected to the Night Hawk* via *reflective memory, enables customers to configure very large systems without impacting the system CPU or VME backplanes with large numbers of programmed input–output transfers.*
3. *This research aims to monitor the continuing but tentative humanitarian, parliamentary and economic contacts between the regimes to detect any changes in their positions, and to gauge what impact their contacts are having upon their domestic politics.*
4. *Thus the research impacted on the team in several important and positive ways.*
5. *This paper will review the range of statutory regulations which impact the means by which data is stored; the conditions under which such storage must occur and the rules regarding the release of such information.*
6. *Univel acknowledged its cuts would impact the Santa Cruz Operation, which it described as a 'partner,' but explained that its focus was on Microsoft.*
7. *Aran doesn't expect the Transaction Point acquisition to impact end of year net profits to March 31 199.*
8. *To ask the Secretary of State for the Home Department what impact recent changes to the conditions of special constables have had on recruiting; and what plans he has further to increase numbers of specials.*
9. *Samsung's original plans were impacted by HP's trouble getting floating point units out of Texas Instruments Inc.*

### Summary and conclusion

This study was conducted to examine potential variations in the collocational behaviors of the near-synonymous verbs *affect* and *impact*. The examination of adverbial and nominal collocations of the two verbs revealed some finite similarities and differences that are not explicated by the dictionary definitions. Both *affect* and *impact* have been shown to prime mainly abstract entities in the subject category and abstract and concrete object collocates. The analysis also revealed some significant differences between the verbs. It was found that *affect* tends to be more frequent with more frequent collocates in the written BNC corpus. It also shows that *affect* as a verb tends to collocate with nominal subjects and adverbs with more forceful and negative connotations. Although Partington (1998) contended that the object of verbs can be indicative of their semantic prosody, the current study findings shows that subject and adverbial verb collocates can also be strong indicators of verbs' favorable semantic connotations. The examination of concordance lines suggests

that *impact* tends to be used in more restricted contexts- primarily in business-technology- than *affect* collocates. In addition, the analysis revealed variations in the preference of syntactic placement of adverbials. While *affect* is more likely to collocate with pre-modifying adverbials, *impact* tends to collocate with post-modifying adverbs.

## Limitations and implications for future studies

The findings of this study can be insightful to ESL teachers in explaining the usage patterns of near-synonymous words, which is usually cited as a common problem among ESL learners. Nevertheless, this study is limited by the number of verbs examined and the domain examined. Future studies are recommended to target a wider set of frequently used near-synonymous verbs. The study findings revealed that the verb *impact* is relatively infrequent in the BNC written corpus, with a relatively small range of collocates. It would be more insightful to examine whether written texts in an American corpus (e.g., COCA) would reveal similar results. Furthermore, examining whether the *impact* noun form displays similar collocational behavioral would offer insightful information on the collocational patterns of the word's different forms. The current findings are limited to written texts, thus future research is recommended to examine potential differences between spoken and written registers and across different disciplines. By studying how near synonyms behave in various disciplines, one hopes that teaching materials can be developed to help ESL learners understand the contextual differences in using near synonyms.

### Applications for ESL pedagogy

The findings of this study offer various applications for both the practical and research realms. The use of corpora to introduce collocational patterns of near synonyms is more informative than providing learners with lists of synonymous words. The collocational patterns can be introduced inductively (i.e., through encouraging learners to pinpoint differences from the summary lists) or deductively, such as when the teacher presents the differences in usage patterns using examples. The variations in colligational preferences between the two near-synonymous verbs in the current study accentuate the pedagogical benefit of drawing learners' attention to the syntactic functions and the type of nouns associated with near-synonymous verbs. Additionally, and considering that part of the difficulty in learning L2 near synonyms is that they usually have one equivalent form in L1, another possible application of the current research is to utilize an explicit contrastive analysis of collocational behaviors of near- synonymous words in L1 versus L2 in teaching near synonyms. The findings can also be used in the field of language learning research to further examine whether collocates of near-synonymous words or collocates with higher LogDice scores or greater frequencies might be processed (or noticed, stored, or retrieved) differently by ESL learners.

## Appendix A: The first 100 subject collocates of affect

| Rank | Subject collocate | Raw frequency | LogDice score |
|---|---|---|---|
| 1 | Factor | 182 | 9.53 |
| 2 | Change | 151 | 8.87 |
| 3 | Decision | 86 | 8.21 |
| 4 | Issue | 76 | 8.11 |
| 5 | Matter | 60 | 7.98 |
| 6 | Recession | 38 | 7.79 |
| 7 | Disease | 41 | 7.72 |
| 8 | Condition | 47 | 7.54 |
| 9 | HIV | 20 | 7.12 |
| 10 | Legislation | 27 | 7.1 |
| 11 | War | 39 | 7.09 |
| 12 | Problem | 56 | 6.94 |
| 13 | Action | 32 | 6.93 |
| 14 | Crisis | 21 | 6.91 |
| 15 | Influence | 20 | 6.87 |
| 16 | Policy | 35 | 6.81 |
| 17 | Event | 28 | 6.78 |
| 18 | Aids | 16 | 6.76 |
| 19 | Law | 34 | 6.72 |
| 20 | Way | 26 | 6.68 |
| 21 | Trend | 16 | 6.57 |
| 22 | Animosity | 13 | 6.55 |
| 23 | Climate | 14 | 6.54 |
| 24 | Pollution | 14 | 6.52 |
| 25 | Development | 23 | 6.52 |
| 26 | Process | 26 | 6.51 |
| 27 | Consideration | 16 | 6.51 |
| 28 | Illness | 14 | 6.49 |
| 29 | Proposal | 18 | 6.42 |
| 30 | Presence | 14 | 6.4 |
| 31 | Activity | 21 | 6.38 |
| 32 | Circumstance | 15 | 6.37 |
| 33 | Variable | 13 | 6.36 |
| 34 | Rule | 20 | 6.34 |
| 35 | Turn | 15 | 6.33 |
| 36 | Strike | 13 | 6.3 |
| 37 | Regulation | 14 | 6.29 |
| 38 | Level | 18 | 6.28 |
| 39 | Drought | 11 | 6.28 |
| 40 | Mutation | 11 | 6.26 |
| 41 | Environment | 14 | 6.25 |
| 42 | Closure | 11 | 6.23 |
| 43 | Cut | 12 | 6.22 |
| 44 | Uncertainty | 11 | 6.2 |
| 45 | Pressure | 14 | 6.2 |
| 46 | Experience | 17 | 6.2 |
| 47 | Shortage | 10 | 6.11 |
| 48 | Measure | 13 | 6.09 |
| 49 | Virus | 10 | 6.06 |
| 50 | Unemployment | 12 | 6.06 |

| Rank | Subject collocate | Raw frequency | LogDice score |
|---|---|---|---|
| 51 | Act | 18 | 6.05 |
| 52 | Death | 14 | 6.03 |
| 53 | Supply | 11 | 6.03 |
| 54 | Gravity | 9 | 5.97 |
| 55 | Context | 10 | 5.97 |
| 56 | Practice | 14 | 5.95 |
| 57 | Covenant | 9 | 5.95 |
| 58 | Constraint | 9 | 5.92 |
| 59 | Loss | 11 | 5.92 |
| 60 | Attitude | 11 | 5.92 |
| 61 | Alcohol | 9 | 5.91 |
| 62 | Fall | 9 | 5.9 |
| 63 | Weather | 11 | 5.89 |
| 64 | Restriction | 9 | 5.85 |
| 65 | Accident | 10 | 5.82 |
| 66 | Erosion | 8 | 5.79 |
| 67 | Provision | 11 | 5.79 |
| 68 | Rate | 15 | 5.77 |
| 69 | Nature | 10 | 5.75 |
| 70 | Injury | 9 | 5.74 |
| 71 | Disorder | 8 | 5.74 |
| 72 | Agreement | 12 | 5.73 |
| 73 | Stress | 8 | 5.7 |
| 74 | Variation | 8 | 5.69 |
| 75 | Movement | 12 | 5.69 |
| 76 | Incident | 9 | 5.69 |
| 77 | Area | 17 | 5.65 |
| 78 | Feeling | 9 | 5.63 |
| 79 | SPR | 7 | 5.62 |
| 80 | Divorce | 7 | 5.58 |
| 81 | Introduction | 7 | 5.58 |
| 82 | Inflation | 8 | 5.57 |
| 83 | Arrangement | 9 | 5.57 |
| 84 | Decline | 7 | 5.54 |
| 85 | Structure | 10 | 5.54 |
| 86 | Difference | 10 | 5.53 |
| 87 | Kind | 8 | 5.52 |
| 88 | Fear | 8 | 5.49 |
| 89 | Move | 8 | 5.46 |
| 90 | Transaction | 7 | 5.45 |
| 91 | Culture | 8 | 5.45 |
| 92 | Neuropathy | 6 | 5.44 |
| 93 | Imposition | 6 | 5.44 |
| 94 | Damp | 6 | 5.43 |
| 95 | Handicap | 6 | 5.39 |
| 96 | Increase | 7 | 5.37 |
| 97 | Treaty | 7 | 5.37 |
| 98 | Flood | 6 | 5.37 |
| 99 | Age | 8 | 5.35 |
| 100 | Shock | 6 | 5.34 |

## Appendix B: The first 100 object collocates of affect

| Rank | Object collocate | Raw freq | LogDice |
|---|---|---|---|
| 1 | Performance | 85 | 7.82 |
| 2 | Life | 145 | 7.78 |
| 3 | Health | 65 | 7.75 |
| 4 | Behaviour | 73 | 7.67 |
| 5 | Area | 129 | 7.66 |
| 6 | Outcome | 60 | 7.63 |
| 7 | People | 171 | 7.12 |
| 8 | Attitude | 44 | 6.98 |
| 9 | Price | 60 | 6.98 |
| 10 | Level | 63 | 6.96 |
| 11 | Ability | 42 | 6.92 |
| 12 | Rate | 59 | 6.87 |
| 13 | Woman | 75 | 6.83 |
| 14 | Industry | 40 | 6.77 |
| 15 | Individual | 36 | 6.74 |
| 16 | Quality | 41 | 6.73 |
| 17 | Relationship | 46 | 6.7 |
| 18 | Trade | 32 | 6.65 |
| 19 | Everyone | 36 | 6.65 |
| 20 | Business | 48 | 6.62 |
| 21 | Output | 30 | 6.6 |
| 22 | Economy | 32 | 6.58 |
| 23 | Decision | 51 | 6.55 |
| 24 | Country | 44 | 6.55 |
| 25 | Family | 39 | 6.51 |
| 26 | Environment | 31 | 6.51 |
| 27 | Result | 48 | 6.51 |
| 28 | Child | 69 | 6.49 |
| 29 | Right | 62 | 6.48 |
| 30 | Balance | 30 | 6.46 |
| 31 | Function | 34 | 6.45 |
| 32 | Relation | 31 | 6.45 |
| 33 | Pattern | 37 | 6.45 |
| 34 | Person | 43 | 6.43 |
| 35 | Structure | 35 | 6.42 |
| 36 | Aspect | 31 | 6.41 |
| 37 | Property | 34 | 6.4 |
| 38 | Development | 36 | 6.38 |
| 39 | Group | 46 | 6.38 |
| 40 | Supply | 27 | 6.34 |
| 41 | Sector | 24 | 6.33 |
| 42 | Whole | 24 | 6.33 |
| 43 | Character | 27 | 6.23 |
| 44 | Community | 25 | 6.22 |
| 45 | Process | 35 | 6.15 |
| 46 | Policy | 35 | 6.12 |
| 47 | Market | 28 | 6.1 |
| 48 | Nature | 24 | 6.06 |
| 49 | Company | 37 | 6.03 |

| Rank | Object collocate | Raw freq | LogDice |
|---|---|---|---|
| 50 | Population | 22 | 6.03 |
| 51 | Operation | 25 | 6.02 |
| 52 | Interest | 37 | 5.99 |
| 53 | Production | 21 | 5.96 |
| 54 | Validity | 17 | 5.95 |
| 55 | Education | 21 | 5.94 |
| 56 | Activity | 27 | 5.94 |
| 57 | Distribution | 18 | 5.94 |
| 58 | State | 25 | 5.92 |
| 59 | Region | 19 | 5.92 |
| 60 | Sale | 21 | 5.91 |
| 61 | Other | 24 | 5.91 |
| 62 | Way | 78 | 5.9 |
| 63 | Value | 29 | 5.88 |
| 64 | System | 43 | 5.87 |
| 65 | Position | 29 | 5.86 |
| 66 | Fish | 20 | 5.85 |
| 67 | Choice | 24 | 5.84 |
| 68 | Demand | 23 | 5.84 |
| 69 | Perception | 16 | 5.83 |
| 70 | Amount | 26 | 5.83 |
| 71 | Interpretation | 17 | 5.83 |
| 72 | Site | 21 | 5.8 |
| 73 | Size | 19 | 5.78 |
| 74 | Response | 21 | 5.78 |
| 75 | Number | 46 | 5.77 |
| 76 | Survival | 15 | 5.77 |
| 77 | Growth | 18 | 5.7 |
| 78 | Minority | 14 | 5.66 |
| 79 | Land | 20 | 5.66 |
| 80 | Use | 26 | 5.62 |
| 81 | Flow | 15 | 5.6 |
| 82 | Work | 41 | 5.6 |
| 83 | Brain | 14 | 5.57 |
| 84 | Variable | 14 | 5.57 |
| 85 | Productivity | 13 | 5.57 |
| 86 | Success | 18 | 5.56 |
| 87 | Service | 29 | 5.56 |
| 88 | Situation | 20 | 5.54 |
| 89 | Practice | 18 | 5.53 |
| 90 | Metabolism | 12 | 5.52 |
| 91 | Worker | 17 | 5.52 |
| 92 | Patient | 19 | 5.51 |
| 93 | Body | 24 | 5.51 |
| 94 | Career | 15 | 5.51 |
| 95 | Employment | 14 | 5.5 |
| 96 | Entitlement | 12 | 5.49 |
| 97 | Employee | 14 | 5.47 |
| 98 | Climate | 12 | 5.45 |
| 99 | Future | 14 | 5.44 |
| 100 | Mood | 12 | 5.41 |

**Abbreviations**
MI: Mutual information index; EFL: English as a foreign language; BNC: British National Corpus; COCA: Corpus of Contemporary American English.

**Acknowledgements**
Not applicable.

**Author's Information**
The author holds a Ph.D. in Applied linguistics and Technology from Iowa State University. She currently works as an assistant professor of English in the College of Education at Majmaa University, KSU. Her major interests include discourse analysis, writing assessment and the use of technology in teaching and learning.

**Authors' contributions**
The author read and approved the final manuscript.

**Funding**
Not applicable.

**Availability of data and materials**
The data used in this work is available for reviewers on request at any time.

## Declarations

**Ethics approval and consent to participate**
Not Applicable.

**Consent for publication**
Not applicable.

**Competing interests**
Not applicable.

Received: 11 December 2021 Accepted: 23 March 2022
Published online: 10 August 2022

**References**

Ackermann, K., & Chen, Y. (2013). Developing the academic collocation list (ACL)—A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*, 235–247. https://doi.org/10.1016/j.jeap.2013.08.002
Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing, 26*, 28–41. https://doi.org/10.1016/j.jslw.2014.09.004
Chan, A. Y. (2010). Toward a taxonomy of written errors: Investigation into the written errors of Hong Kong Cantonese ESL learners. *Tesol Quarterly, 44*(2), 295–319. https://doi.org/10.5054/tq.2010.219941
Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics, 16*(1), 22–29.
Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45–61. https://doi.org/10.1017/S0267190512000074
Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.
Divjak, D. (2006). Ways of intending: Delineating and structuring near synonyms. In S. Gries & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (pp. 19–56). Mouton de Gruyter.
Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157–169.
Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *47*(2), 157–177. https://doi.org/10.1515/iral.2009.007
Dushku, S., & Paek, Y. (2021). Investigating ESL learners' awareness of semantic prosody across proficiency levels. *Language Awareness, 30*(3), 234–256. https://doi.org/10.1080/09658416.2020.1871360
Edmonds, P., & Hirst, G. (2002). Near synonyms and lexical choice. *Computational Linguistics, 28*(2), 105–144.
Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition, 24*(2), 143–188.
Ellis, N., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375–396.
Firth, J. (1957). *Papers in Linguistics*. Oxford University Press.
Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language Learning, 67*(10), 130–154. https://doi.org/10.1111/lang.12226
Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 145–160). Clarendon Press.
Granger, S., & Meunier, F. (2008). Phraseology in language learning and teaching: Where from here? In S. Granger & F. Meunier (Eds.), *Phraseology in foreign language learning and teaching* (pp. 247–252). Binjamins.
Harmon, J. M., Hedrick, W. B., & Fox, E. A. (2000). A content analysis of vocabulary instruction in social studies textbooks for grades 4–8. *The Elementary School Journal, 100*, 253–271.
Hindl, S. (2010). Essential collocations for learners of English: The role of collocational directional weight. In S. Granger & F. Meunier (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 43–66). Binjamins.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.
Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 161–186). Clarendon Press.
Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
Inkpen, D., & Hirst, G. (2006). Building and using a lexical knowledge base of near-synonym differences. *Computational Linguistics, 32*(2), 223–262.
Laufer, B. (1990). Words you know: How they affect the words you learn. In J. Fisiak (Ed.), *Further insights into contrastive linguistics* (pp. 573–593). John Benjamins.
Laufer, B., & Waldman, T. (2011). Verb noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning, 61*(2), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x
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
Liu, D. (2010). Is it a chief, main, major, primary, or principal concern? A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics, 15*(1), 56–87. https://doi.org/10.1075/ijcl.15.1.03liu
Liu, D., & Zhong, S. (2016). L2 vs. L1 use of synonymy: An empirical study of synonym use/acquisition. *Applied Linguistics, 37*(2), 239–261. https://doi.org/10.1093/applin/amu022
Murphy, M. L. (2003). *Semantic relations and the lexicon: Antonymy, synonymy and other paradigms*. Cambridge Univeristy Press.
Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.
Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
Nattinger, J., & DeCarrico, J. (1992). *Lexical phrases and language teaching*. Oxford University Press.
Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*(2), 223–242.
Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research, 21*(3), 298–320. https://doi.org/10.1177/1362168816639619
Osborne, J. (2008). Phraseology effects as a trigger for errors in L2 English. The case of more advanced learners. In S. Granger & F. Meunier (Eds.), *Phraseology in foreign language learning and teaching* (pp. 67–83). Amsterdam: Binjamins.
Paquot, M. (2018). Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly, 15*(1), 29–43. https://doi.org/10.1080/15434303.2017.1405421
Partington, A. (1998). *Patterns and meanings: Using corpora for English language research and teaching*. John Benjamins.
Partington, A., Duguid, A., & Tylor, C. (2013). *Patterns and meanings in discourse: Theory and Practice in corpus-assisted discourse studies (CADS)*. John Benjamins Publishing Company.
Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics, 31*(4), 487–512.
Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.
Sinclair, J. (1998). *The lexical item*. John Binjamins B.V.
Sinclair, J. (2004). The lexical item. In J. Sinclair & R. Carter (Eds.), *Trust the text. Language, corpus and discourse* (pp. 131–148). Routledge.
Stewart, D. (2010). *Semantic prosody: A critical evaluation*. Routledge.
Stubbs, M. (1995). Collocations and semantic profiles. On the cause of the trouble with quantitative studies. *Functions of Language, 2*(1), 23–55.
Stubbs, M. (2013). Sequence and order: The Neo-Firthian tradition of corpus semantics. In H. Hasselgård, J. Ebeling, & S. Oksefjell Ebeling (Eds.), *Corpus perspectives on patterns of lexis* (pp. 13–34). John Benjamins.
Waring, R. (1997). The negative effects of learning words in semantic sets: A replication. *System, 25*, 261–274.
Wible, D. (2010). Multiword expressions and the digital turn. In S. Granger & F. Meunier (Eds.), *Phraseology in foreign language learning and teaching* (pp. 1163–1181). Binjamins.
Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied Linguistics, 27*(1), 103–129. https://doi.org/10.1093/applin/ami045
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

## Publisher's Note