Enhancing EFL Students' COCA-Induced Collocational Usage of Coronavirus: A Corpus-Driven Approach

Amir H.Y. Salama[1]
Department of English
College of Science & Humanities
Prince Sattam Bin Abdulaziz University
Alkharj, Saudi Arabia
Department of English, Faculty of Al-Alsun (Languages)
Kafr El-Sheikh University, Egypt

Waheed M. A. Altohami[2]
Department of English
College of Science & Humanities
Prince Sattam Bin Abdulaziz University
Alkharj, Saudi Arabia
Department of Foreign Languages, Faculty of Education
Mansoura University, Egypt

***Abstract*—The present study seeks to propose a novel pedagogical strategy for enhancing EFL students' collocational usage of the node 'coronavirus' as currently used in the Corpus of Contemporary American English (COCA) across its five genre-based sections, viz. TV/Movies, Blog, Web-General, Spoken, Fiction, Magazine, Newspaper, and Academic. Drawing on a corpus-driven approach, we conducted a pedagogical descriptive analysis of the 'coronavirus' top collocates generated by the COCA. The target collocates have been calculated by the Mutual Information (MI) of 3 or above and specified in terms of the four main lexical parts of speech of nouns, verbs, adjectives, and adverbs. The study has reached three main results. First, employing the COCA as a pedagogical corpus tool can enhance the collocational competence of EFL students should a corpus-driven approach be used descriptively in the classroom. Second, the two methodological stages of demonstration and praxis could facilitate the process of topical priority as a significant index of collocational usage and its thematic relevance. Third, more empirically, the naturally occurring collocates of the node 'coronavirus' have proven significant to the pedagogical situation of teaching the node's collocational meanings encoded in the syntactic categories of nouns, verbs, adjectives, and adverbs, e.g. *infection*, *cause*, *novel*, and *closely*, respectively.**

Keywords—COCA; collocations; coronavirus; corpus-driven approach; EFL learners; extended lexical units

I. Introduction

Corpus-bereft research on collocations and their EFL usage can be said to remain captive of a good deal of misconceptions about the pedagogical nature of teaching and learning vocabulary at large. This type of research can readily be cited in support of the seemingly challenging claim stated above [1, 2, 3-6, 7]. Lamentably, drawing on limited sets of data, research of the sort has offered results about EFL collocational usage that are insensitive to balanced genres in the target learning language, not least because the various patterns of co-occurring words have been conspicuously absent from the analysis of lexical items with collocates used in various domains of human experience.

In an attempt to tout a practical solution to the foregoing problem, we propose to utilize the Corpus of Contemporary American English, commonly known and cited as COCA [8]. The node selected as a model for corpus collocational analysis is the lexical item 'coronavirus' as a currently globally used term in multifarious English-language genres. Towards the collocational analysis of 'coronavirus' in the COCA, we adopt a corpus-driven approach of the extended lexical unit [9, 10] as a methodological tool whereby the following research question can be addressed: How can the COCA be utilized in enhancing EFL students' collocational usage of 'coronavirus'?

Indeed, the question raised above should highlight the significance of the present study as a highly pedagogical and empirical medium for EFL teachers as a community of practice. It is through such a medium that innovative corpus-driven methods can be used for easing EFL students' comprehension of the collocational meanings associated with lexemes of wide-scale thematic relevance and topical interest within the same community of practice. The practical example of the lexeme 'coronavirus', alongside its potential collocates, is claimed to be a practically good and productive site of such relevance and interest. On a more general note, the corpus-driven approach adopted in the current study is likely to secure the ligature between the use of computationally and linguistically tagged corpora such as COCA and the pedagogical applications of teaching one of the most problematic areas in the study of English as an FL – collocational meaning.

In keeping with the corpus-driven approach, therefore, we posit an EFL hypothesis that is subject to empirical validation; the hypothesis can be formulated as such: Applying a corpus-driven approach to the COCA can aid in enhancing the EFL students' collocational usage of the node 'coronavirus'. Addressing the question raised above, and thus (dis-)proving the foregoing hypothesis, the upcoming structure of the paper unfolds in the following way. First, Section 2 presents a review of the literature relevant to the principal point of research. Second, Section 3 elucidates the study's corpus-driven approach as a theoretical framework for the corpus data analysis. Third, Section 4 outlines the current research methodology with a focus on the corpus data (COCA) and the procedure of analysis. Finally, Section 5 concludes the study by offering a discussion of the main findings and a prospect for future research.

II. Literature Review

Language learners' use of collocations received notable scholarly linguistic interest with the focus of exploring their patterns, distribution and frequency, identifying common errors in collocational usage, and experimenting research methods in collocation research. Findings of related studies should ideally be reinforced with investigating the collocational behaviour of pedagogic search terms, identifying their lexico-grammatical patterns, enhancing and testing collocational competence, and gaining insights regarding collocation learning and teaching. The methods employed in collocation research can be generally divided into two directions. The first direction explores the production of collocations by means of large learner corpora [11-18]. The second direction targets collocations collected from questionnaires, interviews, and tests, especially translation [19-22].

Based on collocation sets extracted from learners' essays, Granger [13] compared the way native and non-native learners of English used collocations structured as intensifiers + adjectives. She found that non-native students used atypical word combinations marked as unacceptable by non-native students. Likewise, Howarth [14] focused on the verb-object collocations used by native and non-native learners in different written modes. Findings showed that both native and non-native learners produced non-standard – especially restricted – collocations. Similar findings were reported in Nesselhauf [46] who affirmed that non-native learners' errors in using combinations were distributed over a continuum ranging from free combinations to idioms.

In a similar vein, Durrant and Schmitt [12] compared the native and non-native use of highly frequent collocations used in two parallel corpora composed of students' writing assignments in pre-sessional and in-sessional courses in the UK and Turkey. They targeted manually extracted adjacent modifier-noun word pairs claimed to be particularly common. Using the association measures of Mutual Information (MI) score and T-score, the frequency and strength of such collocations were compared to the same collocations used in BNC as a reference corpus. Findings showed that unlike non-natives, native learners tended to use more low-frequent combinations out of conservatism while writing long essays. Also, non-natives significantly overused strong collocations, but they showed a significant preference to the use of particular combinations.

Altenberg and Granger [11] applied a corpus-based approach, by means of WordSmith Tools, for exploring EFL French/Swedish learners' use of highly frequent collocations based on the verb 'make'. An authentic learner corpus was compared with a native-speaker corpus, namely, the Louvain Corpus of Native English Essays (LOCNESS). Findings highlighted that even advanced learners misused collocations. Although eight uses of 'make'-based collocations have been identified, learners underused delexical (e.g. 'make a decision') and causative (e.g. 'make something possible') structures. A similar approach was followed by Laufer and Waldman [15] who compared learner (the Israeli Learner Corpus of Written English, ILCoWE) and native speaker (LOCNESS) corpora regarding the frequency and correctness of verb-noun collocations. Findings showed that unlike native speakers, non-native learners used fewer collocations. Furthermore, even more advanced learners misused collocations. Also, Paquot and Granger [18] explored the use of English formulaic language in learner corpora, including collocations, phrasal verbs, compounds, idioms, speech formulae, etc. Findings affirmed the relative negative impact of L1 on learners' use of formulaic language regardless of their proficiency level.

Li and Schmitt [23] were concerned with how far L2 learners' collocational competence develops over a year of training on the usage of collocations in an academic writing course in an MA English language teaching program. The reported findings showed no statistically significant development in learners' knowledge of collocations as they tended to overuse specific collocations. Certain errors remained unchanged as learners relied heavily on creativity rather than following lexical patterning. Similarly, Nguyen and Webb [17] explored Vietnamese EFL learners' knowledge of collocations at different frequency levels, the correlation between knowledge of collocations and single-word items, and the predictors of receptive knowledge of collocation. Findings affirmed the positive correlation between knowledge of collocations and single-word items. Also, the major predictors of receptive knowledge – and accordingly the learnability – of collocations included node word frequency, collocation frequency, mutual information score, collocation congruency, and part of speech.

Bahns and Eldaw [19] focused on testing the collocational competence of advanced EFL German learners' by means of translation activities and a cloze test. Findings showed that students sought to paraphrase collocations. Even though some collocations were successfully paraphrased, most paraphrases were unacceptable. Therefore, paraphrasable collocations should not be given prominence in English language teaching. Unlike Bahns and Eldaw [19], Farghal and Obiedat [21] compared the collocational competence of two groups. While the first group included junior and senior Jordanian students at Yarmouk University, the second included English language teachers. Towards this objective, two questionnaires have been administered in the form of fill-in-the-blank and translation tasks. Findings demonstrated that learners' deficiency in using collocations forced them to use lexical simplification strategies, e.g. synonymy, paraphrasing, avoidance, and transfer.

Biskup [20] explored the challenges that faced Polish and German university learners in translating lexical collocations into English. Such translations were then assessed by native speakers of English in terms of acceptability and equivalence. Findings affirmed that both Polish and German students had translational errors. However, while German learners' errors were due to similarity in form, Polish learners' errors were ascribed largely to extending the meaning of L1 collocations to L2. Similarly, Hasselgren [22] explored Norwegian learners' awareness of English collocations during translation tasks. He affirmed that learners' misunderstanding and poor knowledge of collocations led them to rely on literal translation creating what he describes as “collocational dissonance.” That is, though the emerging collocations were grammatically sound, they were not native-like.

Given the Corpus of Contemporary American English (COCA) as the target corpus of present study, several studies have affirmed its efficacy in enriching students’ collocational use especially in writing [24-29]. Hu [25] explored the challenge that near-synonyms impose on the learnability and use of collocations for EFL students. The target synonymous adjective pairs were ‘initial/preliminary’, ‘following/ subsequent’, and ‘sufficient/adequate’. Whilst such pairs were used interchangeably in isolation, findings showed that these collocates designate different prosodies (positive, neutral, and negative) in academic discourse with diverse attitudinal and evaluative meanings.

Mansour [27] sought to foster L2 students’ use of collocations for improving their writing competence and translation performance through getting them to use COCA effectively. Using the COCA’s list display and collocates display options, students have shown significant development in using collocations after receiving the proper training. Likewise, following quasi-experimental research design, Kartal and Yangineksi [26] explored how EFL students learn and produce verb-noun collocations. Hence, experimental and control groups were created, and a collocation knowledge test was administered before and after training students to use collocations through the COCA concordance tool. Findings showed statistically significant differences between the experimental and control groups in terms of the production of collocations. Yet, no significant differences have been noted regarding their collocational knowledge. Similarly, Fang, Ma and Yan [24] explored the way corpus-based training on data-driven learning activities could improve Chinese secondary school students’ writing performance and vocabulary competence in IELTS, including the use of collocations. Students were trained to search two main corpora: COCA and Word and Phrase Concordance. Towards fulfilling this main objective, pre-writing and post-writing tests were used. Findings affirmed that students’ performance in word selection significantly improved as the frequency of collocational errors decreased.

Oktavianti and Sarage [28] studied the frequent and strong collocates of the adjectives ‘great’ and ‘good’ in a corpus compiled from Indonesian EFL textbooks and compared them with those used in COCA. Based on the MI score of collocates, both corpora were similar regarding the verb + adjective structure (e.g. ‘look great/good’). However, considerable mismatches were reported regarding the adverb + adjective structure (e.g. ‘pretty good’ and ‘unpredictably great’), and prominent collocations following the structure of adjective + noun (e.g. ‘great deal’ and ‘good idea’) were markedly absent. Accordingly, textbooks were recommended to be re-examined to render the presented collocations more authentic. Relatedly, Wu [29] investigated how Taiwanese students studying English used the COCA, in an essay writing course, for discovering the collocational patterns of thirty near-synonymous change-of-state verbs. Towards this objective, mixed methods were used including pre-, post-, and delayed post-tests, video files of corpus consultation, a questionnaire, and interviews. Findings showed that although students had some challenges in using the COCA in correcting their miscollocations while drafting their essays, their performance in using collocations improved and such improvement lasted for a considerable time as affirmed by the delayed post-test results.

In view of the foregoing literature review, there seems to be a problematic paucity of corpus-driven investigations of collocations that reflect globally thematic significance and relevance to EFL students/learners in general. Indeed, the collocational use of lexemes whose magnitude of topical saliency and eventfulness is imposing in various semantic domains of expression can be crucial to EFL learners/students at the pedagogical level. One such exemplar is the globally used search-term lemma ‘coronavirus’; and since the term has become a de facto topical attraction in classrooms, either in translation or in writing, there needs to be a particular concern with and focus on the lemma’s collocational usage. This should be especially so at the syntactic level of different parts of speech in genre-balanced corpora wherein collocational usage is likely to be conducive to enhancing competence and developing idiomatic expression. As a corollary of this research gap, the present study attempts to investigate the collocational usage of the lemma ‘coronavirus’ in the COCA in a bid to enhancing the EFL competence of using the currently widely used lemma, and thereby improving the students’/learners’ performance when it comes to using the word-forms associated with this lemma in various pedagogic settings.

III. Theoretical Framework

The lexical meaning of a word is often determined in light of the words that syntagmatically co-occur with it. Such words that tend to hang out together as ready-made chunks came to be known as ‘collocations’. A collocation is commonly viewed as a multi-word formulaic unit (lexical bundle) just like idioms (e.g. ‘back to square one’), proverbs (e.g. ‘let’s make hay while the sun shines’), functional expressions (e.g. ‘excuse me’), fillers (e.g. ‘kind of’), and standardized phrases (e.g. ‘there is a growing body of evidence that’) [30]. Cruse [31] defines collocations as “sequences of lexical items which habitually co-occur” (p. 40). Indeed, Nattinger and DeCarrico [4] define collocations as “strings of words that seem to have certain ‘mutual expectancy’, or a greater-than-chance likelihood that they will co-occur in any text” (p. 21). Hoey [32] affirms that high frequency is the most salient principle marking the behaviour of collocations in a language. He assumes that collocations refer to “the relationship a lexical item has with items that appear with greater than random probability in its (textual) context” (p. 7) [32].

Despite its classification as a unit of formulaic language, a collocation, unlike an idiom, is a compositional phraseme (i.e. a phraseological unit) since its meaning is relatively transparent, i.e. it can – but not necessarily – be inferred from the meaning of its individual parts. Structurally speaking, a collocation is substitutable even when synonyms or near-synonyms are applied; for instance, ‘strong/powerful argument’ is a recurrent collocation, while the collocations 'strong car' and 'powerful tea' are awkward. Also, while a combination such as 'good fortune' is recurrent, the combination 'nice fortune' is semantically unacceptable. Furthermore, unlike free combinations, collocations are somehow considered "grammar in terms of vocabulary" (p. 216) [33], i.e. their co-occurrence always adheres to a set of grammatical principles.

Collocations still, however, could be distributed over a phraseological continuum [34] (Fig. 1) ranging from free combinations (e.g. 'want a car'), to restricted colocations (e.g. 'hold a discussion'), and finally to frozen idioms (e.g. 'sweeten the pill') [14]. The items in free combination are easily replaceable in terms of grammar. Yet, unlike frozen idioms, the meaning of collocations is much more transparent.

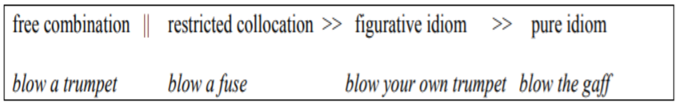


Fig. 1. Cowie's Phraseological Continuum [34].

Gledhill [35] posits that the term 'collocation' tend to signify different notions as far as different perspectives are adopted. First, from a statistical/textual perspective, a collocation signifies a node and its collocates recurrent in a text, i.e. a particular lexical item frequently accompanies another lexical item due to constraints of usage. Therefore, the collocational patterns of a particular phrase are triggered by other phrases at a distance. Second, from a semantic/syntactic level, collocations are approached in terms of lexical combinability, i.e. they are regarded as recurrent, restricted composite units of meaning arranged in particular grammatical sequences, taking into consideration that such sequences are inseparable from their propositional meaning. This collocational restriction means that the meaning of an individual word in specific two-word collocations is restricted or confined to such collocation [36]. For instance, the word 'white' in 'white coffee', 'white noise', and 'white wine' has different senses. Finally, from a discoursal/rhetorical perspective, collocations are assigned diverse pragmatic functions across discourses such as marking topics (e.g. 'let's look at'), shifting topics (e.g. 'ok now'), summarizing (e.g. 'so then'), relating (e.g. 'it has to do with'), and qualifying (e.g. 'the catch is that').

Insofar as the classification of collocations is concerned, two approaches can be enlisted: the phraseological and the distributional [17, 37]. While the phraseological approach focuses on the semantic relation among the words forming the collocation and the non-compositionality of their meaning, the distributional approach focuses on the frequency of a collocation in a corpus or corpora. In view of this demarcation and based on the items forming collocations, collocations are classified into lexical and grammatical collocations. Grammatical collocations are more frequent in English, and they are lexicalized as single units with formulaic meanings. Bahns [38] affirms that grammatical collocations take the form of a noun, an adjective, or a verb followed by either a particle, an infinitive, or a clause, e.g. 'by accident', 'angry at', 'afraid that', and 'adhere to'. Unlike grammatical collocations, lexical collocations have no grammatical elements as they are composed of open-class lexical items [39]. They are structured as noun + noun (e.g. 'ceasefire agreement'), adjective + noun (e.g. 'strong tea'), noun + verb (e.g. 'results showed'), verb + noun (e.g. 'make a mistake'), verb + adverb (e.g. 'walk slowly'), and adverb + adjective (e.g. 'amazingly gorgeous'). Furthermore, clusters of lexical collocations are claimed to share a similar semantic prosody [40]. Yet, based on the frequency of collocations, Hill [41] divides collocations into four types: weak collocations (e.g. 'red wine'), medium-strength collocations (e.g. 'Sun reader'), strong collocations (e.g. 'rancid butter'), and unique collocations (e.g. 'leg room').

With regard to collocational use, a set of parameters have been proposed. One crucial parameter of collocational use is high frequency, i.e. highly frequent word combinations are systematically classified as collocations. Another important parameter is that of word association strength as specific words tend to co-occur biasedly [42]. In this regard, diverse statistical measures could be employed such as Mutual Information (MI), T-Score, and Z-Score. A third parameter is that of substitutability, i.e. how far an item in a collocation could be substituted by a synonym or a near synonym. As an integrative part of any language, collocations are processed in the mind of language user in two different ways: analytic processing and holistic processing [43]. On the one hand, in analytic processing, the lexical items and grammatical patterns of word combinations are computed and then their meanings are retrieved by assembling each item's meaning. This kind of processing occurs at a slower speed with much processing load. Holistic processing, on the other hand, is conducted at a faster speech with less processing load as word combinations are memorized as units (prefabricated forms) whose meanings are relatively difficult to be retrieved from the meanings of their individual parts, such as phrasal verbs and compounds. Still, collocation processing might occur in a parallel mode [43].

Indeed, the surge of publications in collocation teaching and learning is by all accounts an index of the significance of collocation competence, which is in turn crucial to language proficiency. For instance, Gledhill [35] asserts that "it is impossible for a writer to be fluent without a thorough knowledge of the phraseology of the particular field he or she is writing in" (p. 1). Equally important, Hill and Lewis [44] regard collocation use as "one of the most powerful forces in making language coherent, fluent, comprehensible, and predictable" (p. 1). Similarly, in any text, collocations are claimed to have "a cohesive force" [45]. Additionally, Fillmore, Kay, and O'Connor [46] point out that collocations should be integrated in language learning since they are culturally salient.

Hill [41] affirms that collocations represent "the most powerful force in the creation and comprehension of all naturally occurring text" (p. 49). In the context of English Language Teaching (ELT), it has been largely claimed that the accurate use of collocation is an essential component of communicative competence [47] and an indicator of proficiency as collocational knowledge allows native-like language use [48]. That is why the misuse of collocations is envisaged as "a major indicator of foreignness" (p. 232) [38].

That is, most of the collocational errors are experienced by non-native speakers usually due to lack of lexical proficiency. In this regard, Nation [3] explains that less proficient learners tend to "encode words in memory on the basis of sound and spelling rather than by association meaning" (p.3). Similarly, Laufer [49] and Erman, Lundell and Lewis [50] affirm that collocations, among other formulaic units are linked to native speakers' fluent and natural language production as well as linguistic diversity.

Crucially, a methodological distinction is always made between corpus-based and corpus-driven approaches. The corpus-based approach (CBA) targets previously identified linguistic features and constructs as well as patterns of variation and use. In other words, the corpus would support intuitive knowledge, confirm linguistic pre-set assumptions, and provide illustrative examples. Meanwhile, the corpus-driven approach (CDA) aims at exploiting the potential of corpora for the identification of recurrent linguistic categories and patterns emerging in context not fully recognized before [51-52]. Furthermore, CBA starts with no prior assumption, and all conclusions are usually reached relying on corpus observations. The corpus-driven approach has been largely used in the analysis of multi-word sequences known as 'lexical bundles' including idioms, proverbs, and collocations. The target is always their frequency, and distribution, usually followed by an analysis of emerging patterns and functional characteristics.

Indeed, many factors have been reported to affect the learnability of collocations. One of these factors is in the semantic complexity of a collocation. Regarding semantic complexity, collocations could be distributed over a continuum from total transparency to opacity. Figuring out the meaning of a collocation depends on the language user's familiarity with the individual words forming the collocation. Accordingly, it is largely claimed that learners are expected to spot the meaning of free combinations (e.g. 'pay money') more than restricted (e.g. 'pay attention'), and idiomatic collocations (e.g. 'pay lip service') [15, 53]. Furthermore, collocational congruency is also claimed to affect collocational usage as EFL students are reported to make more errors and react more slowly to incongruent collocations than congruent collocations (p. 647) [54]. Specifically in restricted collocations, L2 learners are reported to make "overliberal assumptions about the collocational equivalence of semantically similar items" (p. 202) [48]. That is, they tend to be able to produce atypical word combinations using items with similar meanings, e.g. 'plastic operation' instead of 'plastic surgery'. The reason is that they perceive lexical items individually rather than in combination, and therefore they strategically tend to simplify the lexemes form collocations through the use of synonyms, paraphrasing, and transferring L1 items to L2 through literal translation [21]. Generally, L1 interference is largely claimed to produce many of the errors in collocational usage, even on the part of advanced learners [55]. Such words that learners learnt at early stages and tend to cling to them even after training came to be known as "lexical teddy bears" [22].

Indeed, the introduction of corpus tools and techniques formed a turning point in phraseology studies in general and the study of formulaic units (including collocations) in particular. L2 research benefited greatly from such tools and techniques which offered more comprehensive empirical techniques for building, analyzing, and comparing corpora, thereby allowing the exploration of authentic language as practised by language learners compared to native speakers. This line of research is referred to as 'data-driven learning' (DDL) [9, 56, 57]. In DDL, language is viewed as data and the main objective of DDL tasks is to lead learners to identify patterns and uses of language by means of corpus tools, and thereby developing their autonomy. Software packages such as WordSmith, ConcGram, AntConc, etc. offer tools for calculating frequencies of words and their token/type ratio, extracting concordance lines with the target key words, featuring their various co-texts. As mentioned in the literature review section, corpora – taking native-speaker corpora as the norm – have been employed in investigating the typology of collocations as well as collocational underuse, overuse, and misuse.

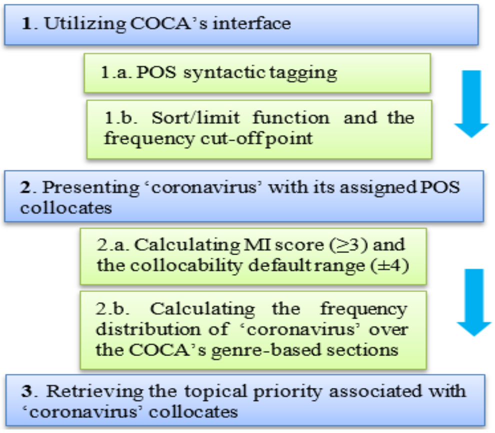
IV. Methodology

A. Data

The present study is geared towards eliciting the highly frequent collocates used by native speakers of General American English (GA) when covering information of different types on the outbreak and progression of 'coronavirus' (or more technically, COVID-19) as represented in the American Corpus of Contemporary American English (COCA) [10]. The COCA has been selected in this study for a host of reasons. First, updated in 2021, the COCA (available at https://www.english-corpora.org/coca/) contains more than one billion words of data distributed over 485,202 texts, and therefore it is regarded as the most widely used, freely available corpus worldwide. Secondly, it is genre-balanced as it offers data covering a wide range of spoken and written, formal and informal genres. These genres are web genres and blog, newspapers, magazines, spoken, academic, fiction, and TV/movies. Thirdly, and finally, the user-friendly interface of COCA allows getting information about – and comparing – the frequency, currency, time span, and prosody of words, phrases, and grammatical constructions across diverse genres. Besides, the COCA offers information on definitions, keyness, related topics, collocates, clusters, lemmas, synonyms, and customized word lists.

B. Procedure

The methodological procedure adopted in the present study is a two-stage process of addressing the primary research question of how the COCA can be utilized in enhancing EFL students' collocational usage of 'coronavirus' (see Fig. 2). The first stage is concerned with setting the pedagogical scene; it amounted to a demonstration of how the COCA's interface can potentially be utilized in terms of its available POS syntactic tagging and the sort/limit function as well as the frequency cut-off point and hits specified. In respect of the second stage of the procedure, the lexical item 'coronavirus' has been presented with its assigned part-of-speech collocates with the automatic generation calculated by an MI score of 3 or above and the collocability default range ±4. At this stage, too, the frequency distribution of 'coronavirus' over the COCA's genre-based sections was calculated based on the generated collocates themselves, and thereafter the topical priority associated with 'coronavirus' collocates in COCA was automatically retrieved in relation to the extracted collocational pairs.



1. Utilizing COCA's interface

1.a. POS syntactic tagging

1.b. Sort/limit function and the frequency cut-off point

2. Presenting 'coronavirus' with its assigned POS collocates

2.a. Calculating MI score (≥3) and the collocability default range (±4)

2.b. Calculating the frequency distribution of 'coronavirus' over the COCA's genre-based sections

3. Retrieving the topical priority associated with 'coronavirus' collocates

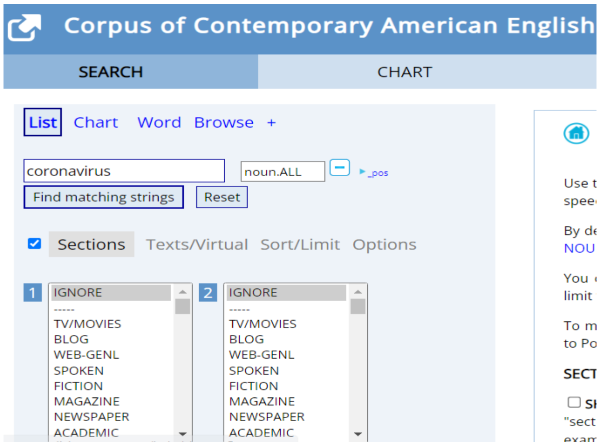
Fig. 2. The Procedure of Analysis.

V. Data Analysis and Discussion

The present section of analysis is divided into two stages. The first stage is dedicated to the demonstration of the COCA's interface as a way of setting the pedagogical scene; the second stage is a proposed EFL pedagogical praxis whereby the automatic generation of 'coronavirus' collocates and their relevant topical priority across the COCA's different genre-specific sections.

A. Setting the Pedagogical Scene: The COCA in Focus

The current stage of analysis can best be described as an EFL demonstration of the COCA's interface. As exhibited in Fig. 3, the lexical item 'coronavirus' has been entered into the COCA's search box with the particular POS tagging noun. ALL, which restricts the search hits to coronavirus as an exclusively nominal form. Also, demonstrably, the different COCA sections are presented for a potential selection, whereby an EFL teacher can decidedly opt for a genre-based search domain for 'coronavirus', say, TV/MOVIES or FICTION; and, perhaps, the teacher can interestingly compare such sections.



coronavirus

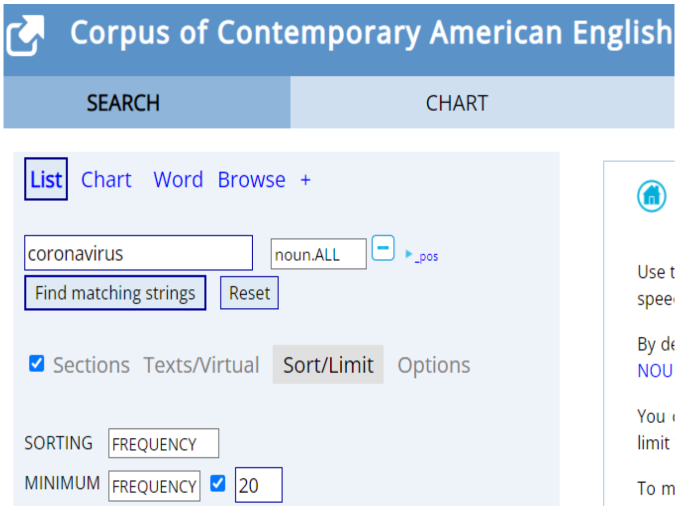
noun.ALL

☑ Sections Texts/Virtual Sort/Limit Options

1	2
IGNORE ----- TV/MOVIES BLOG WEB-GENL SPOKEN FICTION MAGAZINE NEWSPAPER ACADEMIC	IGNORE ----- TV/MOVIES BLOG WEB-GENL SPOKEN FICTION MAGAZINE NEWSPAPER ACADEMIC

Use t
spee
By de
NOU
You
limit
To m
to Po
SECT
□ SI
"sect

Fig. 3. The COCA's Genre-based Sections and POS Tagging.



coronavirus

noun.ALL

☑ Sections Texts/Virtual Sort/Limit Options

SORTING FREQUENCY

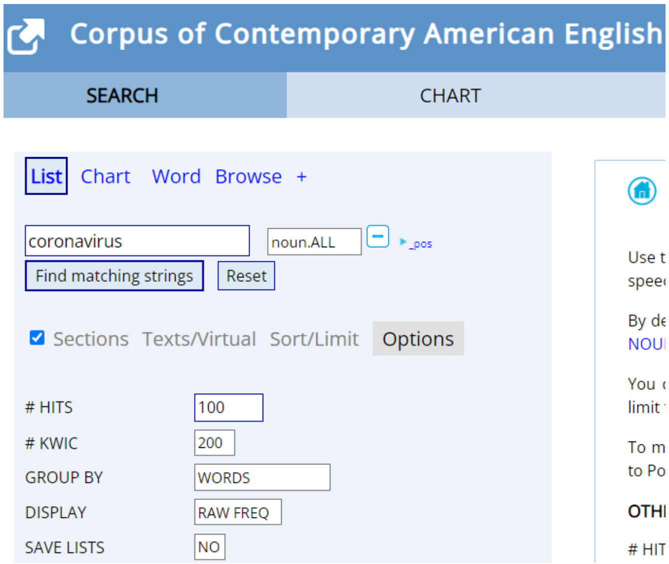
MINIMUM FREQUENCY ☑ 20

Use t
spee
By de
NOU
You
limit
To m

Fig. 4. The COCA's Sort/Limit Function for Lemma Search.

Moving to Fig. 4, EFL students can be trained in how to use the COCA's Sort/Limit function for lemma search. The teacher is just supposed to employ this function as a way of specifying the frequency cut-off point of 20. The function is crucial since it facilitates the pedagogic situation by rendering the search process manageable enough to the students, let alone the fact that the same function generates the highly frequent occurrence of the lexical item as a lemma.

There are yet other COCA-built functions for lemma search as shown in Fig. 5, where other options are displayed. At this point, the teacher should ideally keep the students focused on the number of hits germane to 'coronavirus' (100 times) and the KWIC scope allowed (by default 200) as well as the featured raw frequency; these more options can further be used to facilitate the process of searching for the significant instances of the lemma 'coronavirus'. Of course, thus far, we have not touched upon the actual distribution the lemma ('coronavirus') over the genre-based sections – which is so pedagogically crucial to the recognition of vitiations in use.



coronavirus

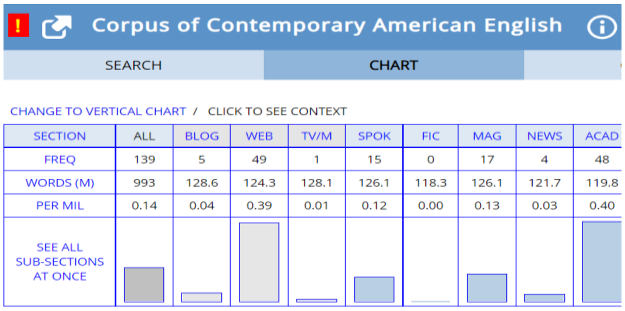
noun.ALL

☑ Sections Texts/Virtual Sort/Limit Options

# HITS	100
# KWIC	200
GROUP BY	WORDS
DISPLAY	RAW FREQ
SAVE LISTS	NO

Use t
spee
By de
NOU
You
limit
To m
to Po
OTHI
HIT

Fig. 5. The COCA's Other Significant Functions for Limiting Lemma Search.


Corpus of Contemporary American English
SEARCH
CHART
CHANGE TO VERTICAL CHART / CLICK TO SEE CONTEXT

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
FREQ	139	5	49	1	15	0	17	4	48
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8
PER MIL	0.14	0.04	0.39	0.01	0.12	0.00	0.13	0.03	0.40
SEE ALL SUB-SECTIONS AT ONCE									

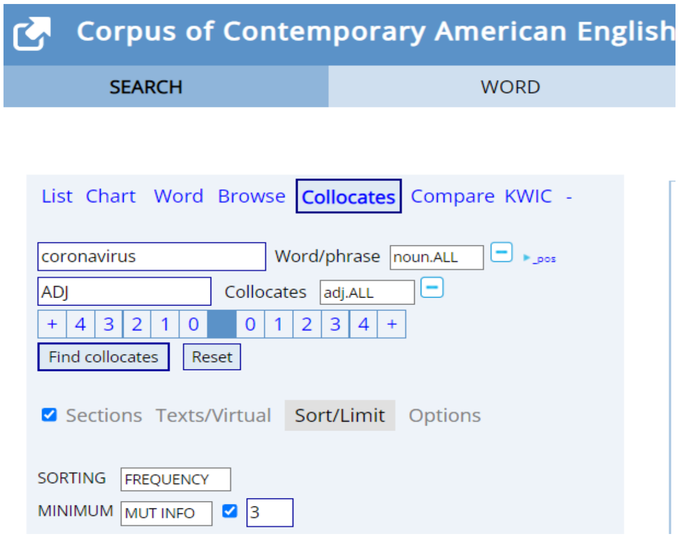
Fig. 6. Frequency Distribution of 'Coronavirus' Over the COCA's Genre-based Sections.

Coming to Fig. 6, the teacher can be said to be able to prepare his/her students for the recognition of the variations in lemma use referred to above, with a closer eye on frequency distribution. Thus, on closer inspection, students will readily observe that 'coronavirus' is most frequently used in the section WEB PAGES next to which in frequency is ACADEMIC. Perhaps, this may be ascribed to the fact that the Internet as an electronic medium has consistently demonstrated global-scale impact due to "its intensity of use" (p. 5) [58]. The frequency associated with the COCA's section ACADEMIC can be explained on the grounds that the topical nature of 'coronavirus' is scientific in the first place. Also, the low frequency of 'coronavirus' in the sections of FICTION (0 frequency), TV/M (1 time), NEWS (4 times), and BLOG (5 times) can be explained against the background genre nature. A good teacher, we argue, ought to think of the widespread use of a given lemma in a certain genre, and here there lies the rub: compared to the WEB PAGES and ACADEMIC, the rest of the low-frequency genres lack in sub-genres. Thus, when it comes to searching for a currently global term such as 'coronavirus', students should be directed to sub-genre-composed genres.

As the current EFL demonstration proceeds, we need to draw teachers' attention to the fact that in order for students to gain the utmost pedagogical benefits out of the COCA's interface, the frequency of 'coronavirus' or of any other search term is far from enough; there needs to be an investigation of the patterns of use associated with 'coronavirus', or again with any comparable term. At this point of analysis, therefore, collocational usage of 'coronavirus' is presented as a pedagogical praxis in the coming sub-section.

B. *Proposing a Pedagogical Praxis: The COCA-Driven Collocates of 'Coronavirus'*

EFL students' competence for the collocational usage of 'coronavirus' – and indeed any other lexical item – can be well enhanced should the teacher make a point of searching for the collocates strongly co-occurring with the node word 'coronavirus'. As exhibited in Fig. 7, this is feasible via the COCA's interface by clicking the 'collocates' icon after specifying the POS tagging of target collocates. The figure features the tagging adj. ALL of such collocates and offers the default range ±4, i.e. four collocates to the right and/or the left of the search term as the topmost span (4:4 collocates). Further, crucially, the teacher needs to present students with the collocation measurement of Mutual Information (MI of 3 or above).



coronavirus Word/phrase noun.ALL

ADJ Collocates adj.ALL

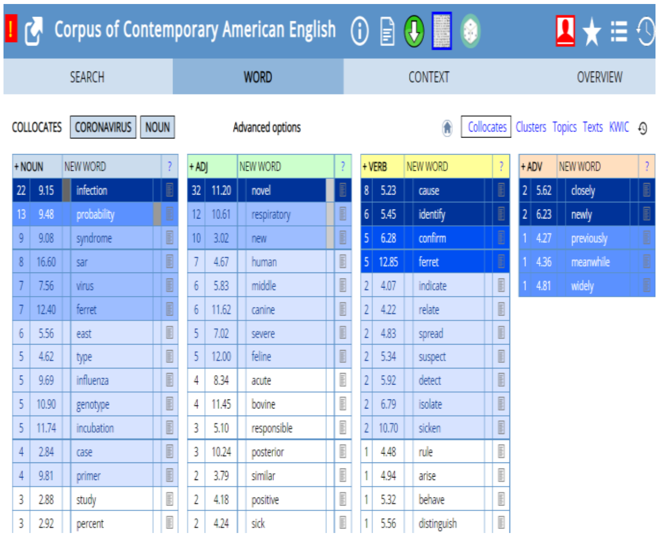
☑ Sections

SORTING FREQUENCY

MINIMUM MUT INFO ☑ 3

Fig. 7. Searching for the Collocates of 'Coronavirus' on COCA's Interface.

Moving to the next practical step in EFL class demands a pedagogically direct engagement with the top collocates of 'coronavirus' in terms of their lexical parts of speech, i.e. nouns, adjectives, verbs, and adverbs. Indeed, this is one of the most useful tools of the COCA's interface. As shown in Fig. 8, the tool is corpus-driven, in that it provides the top part-of-speech collocates based on their frequency of co-occurrence with the node 'coronavirus' in varying degrees of highlighting. Thus, the order of noun collocates are infection, probability, syndrome, sar, virus, and ferret; the adjective collocates are novel, respiratory, and new; the verb collocates are cause, identify, confirm, and ferret; the adverb collocates are closely, newly, previously, meanwhile, and widely. Given such a corpus-driven set, students should be equipped with a whole profile of the main collocates in their various lexical parts of speech; at this point, the teacher is required to engage with the teaching situation by asking students to use collocations in different lexico-grammatical patterns, or to emulate certain native-like usages of comparable patterns.



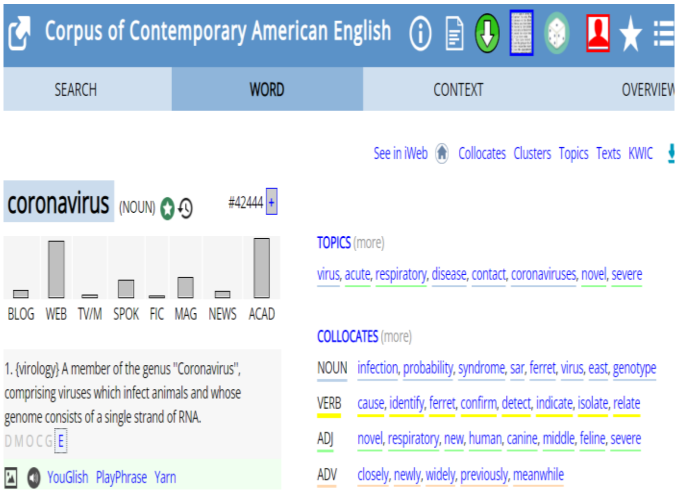
+ NOUN		NEW WORD	?
22	9.15	infection	
13	9.48	probability	
9	9.08	syndrome	
8	16.60	sar	
7	7.56	virus	
7	12.40	ferret	
6	5.56	east	
5	4.62	type	
5	9.69	influenza	
5	10.90	genotype	
5	11.74	incubation	
4	2.84	case	
4	9.81	primer	
3	2.88	study	
3	2.92	percent	

+ ADJ		NEW WORD	?
32	11.20	novel	
12	10.61	respiratory	
10	3.02	new	
7	4.67	human	
6	5.83	middle	
6	11.62	canine	
5	7.02	severe	
5	12.00	feline	
4	8.34	acute	
4	11.45	bovine	
3	5.10	responsible	
3	10.24	posterior	
2	3.79	similar	
2	4.18	positive	
2	4.24	sick	

+ VERB		NEW WORD	?
8	5.23	cause	
6	5.45	identify	
5	6.28	confirm	
5	12.85	ferret	
2	4.07	indicate	
2	4.22	relate	
2	4.83	spread	
2	5.34	suspect	
2	5.92	detect	
2	6.79	isolate	
2	10.70	sicken	
1	4.48	rule	
1	4.94	arise	
1	5.32	behave	
1	5.56	distinguish	

+ ADV		NEW WORD	?
2	5.62	closely	
2	6.23	newly	
1	4.27	previously	
1	4.36	meanwhile	
1	4.81	widely	

Fig. 8. Top Collocates of 'Coronavirus' and their Lexical Parts of Speech in COCA.



1. {virology} A member of the genus "Coronavirus", comprising viruses which infect animals and whose genome consists of a single strand of RNA.

TOPICS (more)

virus, acute, respiratory, disease, contact, coronaviruses, novel, severe

COLLOCATES (more)

NOUN	infection, probability, syndrome, sar, ferret, virus, east, genotype
VERB	cause, identify, ferret, confirm, detect, indicate, isolate, relate
ADJ	novel, respiratory, new, human, canine, middle, feline, severe
ADV	closely, newly, widely, previously, meanwhile

Fig. 9. Topical Priority Associated with ‘Coronavirus’ Collocates in COCA.

As well as identifying the lexical part-of-speech collocates of ‘coronavirus’, the COCA’s interface has the pedagogically effective feature of what we prefer to call the corpus-driven topical priority associated with the collocates. As demonstrated in Fig. 9, the topical priority generated by the COCA and regarded as thematically relevant to the collocational pairs identified in Fig. 7 consists largely in specific topical domains: virus, acute, respiratory, disease, contact, coronaviruses, novel, and severe. Further, as presented in Fig. 8, the node term ‘coronavirus’ is defined within the topical scope of {virology}, which can be said to reveal the semantic nature of ‘coronavirus’ as [+viral].

Thus, bringing together the last two steps of part-of-speech-bound collocates and their topical priority may well improve the EFL students’ understanding of ‘coronavirus’ as a concept; that is, beyond the term as a de-contextualized lexical item that is isolated from its significant collocates.

VI. Conclusion and Future Research

In conclusion, we are in a position to round off the pedagogical strategy proposed in the present study for enhancing EFL students’ COCA-induced collocational usage of ‘coronavirus’. The approach used towards the fulfilment of this goal has been presented as more corpus-driven than corpus-based. The COCA has been utilized for empirically validating the proposed strategy. Such a strategy can be said to have yielded three results. First, a node word can be semantically defined in relation to its potential collocates provided there should be (a) a lexically orientated part-of-speech framing of these collocates and (b) a genre-sensitive balanced set of data manipulated by corpus software. The present case in point was presented in the nominal form ‘coronavirus’ whose lexical collocates were statistically calculated and formally recognized as nouns, adjectives, verbs, and adverbs in the COCA.

Second, a two-stage investigation of the ‘coronavirus’ node-collocate relation has been undertaken in a pedagogically systematic fashion. The first stage was a demonstration of the COCA’s interface and its main functions of sorting and limiting the searches for a particular term via grammatically annotated settings of POS tagging as well as other relevant functions of specifying frequency and MI collocation statistics. The second stage was provided as a pedagogical praxis with specific highlights: (i) setting the collocation default range ±4, (ii) constructing coronavirus frequency distribution over the COCA’s genre-based sections (TV/Movies, Blog, Web-General, Spoken, Fiction, Magazine, Newspaper, and Academic), (iii) generating the top collocates of ‘coronavirus’ and their lexical parts of speech in the COCA, and bringing out the topical priority associated with ‘coronavirus’ collocates in the same corpus data.

Third, on a rather empirical level, the actual collocates of the node word ‘coronavirus’ have been generated from the COCA as nouns, e.g. infection, probability, syndrome, sar, virus, and ferret; adjectives, e.g. novel, respiratory, and new; verbs, e.g. cause, identify, confirm, and ferret; and adverbs, e.g. closely, newly, previously, meanwhile, and widely. Further, on the same empirical level, collocation-induced topical priority was derived from the COCA in thematic relevance to the above collocates of ‘coronavirus’; and they consisted in the following topical domains: virus, acute, respiratory, disease, contact, coronaviruses, novel, and severe. In view of such an empirical finding, with recurrent COCA generation of ‘virus’, the node ‘coronavirus’ has (perhaps unsurprisingly) been demonstrated to fall in the topical scope of {virology}.

References

[1] H. Lien, “The effects of collocation instruction on the reading comprehension of Taiwan college students,” Unpublished doctoral dissertation, Indiana University of Pennsylvania, Pennsylvania, 2003.

[2] M. J. McCarthy, “A new look at vocabulary in EFL,” *Applied Linguistics*, vol. 5, no. 1, pp. 12-22, 1984.

[3] I. S. Nation, *Teaching and learning vocabulary*. Boston: Heinle & Heinle Publishers, 1990.

[4] I. S. Nation, *Learning vocabulary in another language* (3rd ed.). Cambridge: Cambridge University Press, 2002.

[5] J. R. Nattinger, and J. S. DeCarrico, *Lexical phrases and language teaching*. Oxford: Oxford University Press, 1992.

[6] J. R. Nattinger, and J. S. DeCarrico, *Lexical phrases and language teaching* (2nd ed.). Oxford: Oxford University Press, 1997.

[7] S. Shih, and H. Wang, “The Relationship Between EFL Learners’ Depth of Vocabulary Knowledge and Oral Collocation Errors,” In proceedings of The 23rd International Conference on English Teaching and Learning in the Republic of China, pp. 964-977. Taipei, Taiwan: Kaun Tang International Publishing Ltd., 2006.

[8] M. Davies, The Corpus of Contemporary American English (COCA), Available online at https://www.english-corpora.org/coca/,2008-

[9] J. Sinclair, *Corpus concordance collocation*. Oxford: Oxford University Press, 1991.

[10] J. Sinclair, *Trust the text: Language, corpus, and discourse*. London, New York: Routledge, 2004.

[11] B. Altenberg, and S. Granger, “The grammatical and lexical patterning of MAKE in native and non-native student writing,” *Applied Linguistics*, vol. 22, pp. 173-195, 2001. https://doi.org/10.1093/applin/22.2.173.

[12] P. Durrant, and N. Schmitt, N., “To what extent do native and non-native writers make use of collocations?,” *IRAL-International Review of Applied Linguistics in Language Teaching*, vol. 47, pp. 157-177, 2009. doi:10.1515/iral.2009.007.

[13] S. Granger, “Prefabricated patterns in advanced EFL writing: Collocations and formulae,” In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*, pp. 79-100. Oxford: Oxford University Press, 1998.

[14] P. Howarth, “The phraseology of learners’ academic writing,” In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications,* pp. 161-186. Oxford: Oxford University Press, 1998.

[15] B. Laufer, and T. Waldman, “Verb–noun collocations in second language writing: A corpus analysis of learners,” *English Language Learning*, vol. 61, pp. 647–672, 2011.

[16] N. Nesselhauf, “The use of collocations by advanced learners of English and some implications for teaching,” *Applied Linguistics*, vol. 24, pp. 223-242, 2003.

[17] T. M. H. Nguyen, and S. Webb, “Examining second language receptive knowledge of collocation and factors that affect learning,” *Language Teaching Research*, pp. 1-23, 2016. doi:10.1177/1362168816639619.

[18] M. Paquot, and S. Granger, “Formulaic language in learner corpora,” *Annual Review of Applied Linguistics*, vol. 32, pp. 130-149, 2012. doi:10.1017/S0267190512000098.

[19] J. Bahns, and M. Eldaw, “Should we teach EFL students collocations?,” *System*, vol. 21, pp. 101-114, 1993.

[20] D. Biskup, “L1 influence on learners’ rendering of English collocations: A Polish/German empirical study,” In P. J. L. Arnaud, and H. Bejoint (Eds.), *Vocabulary and Applied Linguistics*, pp. 85-93. London: Macmillan, 1992.

[21] M. Farghal, and H. Obiedat, “Collocations: A neglected variable in EFL,” *International Review of Applied Linguistics*, vol. 33, pp. 315-331, 1995.

[22] A. Hasselgren, “Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary,” *International Journal of Applied Linguistics*, vol. 4, pp. 237–258, 1994. https://doi.org/10.1111/j.1473-4192.1994.tb00065.x.

[23] J. Li, & N. Schmitt, “The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach,” In D. Wood (Ed.), Perspectives on Formulaic Language: Acquisition and Communication, pp. 22-46. New York: Continuum, 2010.

[24] L. Fang, Q. Ma, and J. Yan, “The effectiveness of corpus-based training on collocation use in L2 writing for Chinese senior secondary school students,” *Journal of China Computer-Assisted Language Learning*, vol. 1, no. 1, pp. 80-109, 2021. https://doi.org/10.1515/jccall-2021-2004.

[25] H. C. M. Hu, “A semantic prosody analysis of three adjective synonymous pairs in COCA,” *Journal of Language and Linguistic Studies*, vol. 11, no. 2, pp. 117-131, 2015.

[26] G. Kartal, and G. Yangineksi, “The effects of using corpus tools on EFL student teachers' learning and production of verb-noun collocations,” *PASAA*, vol. 55, pp. 100-122, 2018.

[27] D. M. Mansour, “Using COCA to Foster Students’ Use of English Collocations in Academic Writing,” In proceedings of the 3rd International Conference on Higher Education Advances, HEAd’17 Universitat Politecnica de Valencia, Valencia, pp. 600-607, 2017. DOI: http://dx.doi.org/10.4995/HEAd17.2017.5301.

[28] I. N. Oktavianti, and J. Sarage, “Collocates of ‘great’ and ‘good’ in the corpus of contemporary American English and Indonesian EFL textbooks,” *Studies in English Language and Education*, vol. 8, no. 2, pp. 457-478, 2021. https://doi.org/10.24815/siele.v8i2.18594.

[29] Y. Wu, “Discovering collocations via data-driven learning in L2 writing,” *Language Learning & Technology*, vol. 25, no. 2, pp. 192-214, 2021.

[30] F. Boers, J. Eyckmans, H. Kappel, H. Stengers, & M. Demecheleer, “Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test,” *Language Teaching Research*, vol. 10, pp. 245-261, 2006. https://doi.org/10.1191/1362168806lr195oa.

[31] D. A. Cruse, *Lexical semantics*. New York: Cambridge University Press, 1986.

[32] M. Hoey, *Patterns of lexis in text*. Oxford: Oxford University Press, 1991.

[33] G. Kennedy, “Collocations: Where grammar and vocabulary teaching meet,” *Language Teaching Methodology for the Nineties*, RELC, Anthology Series 24, 1990.

[34] A. P. Cowie, “The treatment of collocations and idioms in learners’ dictionaries,” *Applied Linguistics*, vol. 2, no. 3, pp. 223-235, 1981.

[35] C. J. Gledhill, *Collocation in science writing*. Tubingen: Gunter Narr Verlag, 2000.

[36] D. A. Cruise, “Language, meaning and sense: Semantics. In N. E. (Ed.), *An Encyclopedia of Language*, pp. 139-172. New York: Routledge, 1990.

[37] D. Gablasova, V. Brezina, and T. McEnery, “Collocations in corpus-based language learning research: Identifying, comparing, and interpreting the evidence,” *Language Learning, A Journal of Research in Language Studies*, vol. 67, no. 1, pp. 155-179, 2017. https://doi.org/10.1111/lang.12225

[38] J. Bahns, “Lexical collocations: A contrastive view,” *ELT Journal*, vol. 47, no. 1, pp. 56-63, 1993.

[39] T. Fontenelle, “Lexical functions in dictionary entries. In A. P. Cowie, Phraseology: theory, analysis, and applications, pp. 189-207. Oxford: Oxford University Press, 1998.

[40] W. E. Louw, “Irony in the Text or Insincerity in the Writer?,” In M. Baker (Ed.), *The Diagnostic Potential of Semantic Prosodies.” Text and Technology: In Honour of John Sinclair*, pp. 157-176. Amsterdam: John Benjamins, 1993.

[41] J. Hill, “Revising priorities: From grammatical failure to collocational success,” In M. Lewis, *Teaching Collocations: Further developments in the lexical approach*, pp. 47-69. Hove: Language Teaching Publications, 2000.

[42] S. Hunston, *Corpora in applied linguistics*. Cambridge: Cambridge University Press, 2002.

[43] K. Matsuno, “Processing collocations: Do native speakers and second language learners simultaneously access prefabricated patterns and each single word?,” *Journal of the European Second Language Association*, vol. 1, no. 1, pp. 61–72, 2017. DOI: https://doi.org/10.22599/jesla.17.

[44] J. Hill**,** and M. Lewis (Eds.), *LTP dictionary of selected collocations*. Hove: Language Teaching Publications, 1997.

[45] M. A. K. Halliday, and R. Hasan (Eds.), *Cohesion in English*. Essex: Longman, 1976.

[46] C. J. Fillmore, P. Kay, and M. C. O’Connor, “Regularity and idiomaticity in grammatical constructions: The case of letalone,” *Language*, vol. 64, pp. 501-538, 1988. https://doi.org/10.2307/414531.

[47] M. Stubbs, *Words and phrases*. Oxford: Blackwell, 2001.

[48] A. Wray, *Formulaic language and the lexicon*. Cambridge, England: Cambridge University Press, 2002.

[49] B. Laufer, “The influence of L2 on L1 collocational knowledge and on L1 lexical diversity in free written expression,” In V. Cook (Ed.), *Effects of the Second Language on the First*, pp. 120-141. Clevedon: Cromwell Press Ltd, 2003.

[50] B. Erman, F. Forsberg Lundell, M. Lewis, “Formulaic language in advanced second language acquisition and use,” In K. Hyltenstam (Ed.), *Advanced Proficiency and Exceptional Ability in Second Languages*, pp. 111-148. Boston: Walter de Gruyter, 2016.

[51] C. F. Meyer, “Corpus-based and corpus-driven approaches to linguistic analysis: One and the same?,” In I. Taavitsainen, M. Kytö, C. Claridge, and J. Smith (Eds.), *Developments in English: Expanding Electronic Evidence*. Cambridge: Cambridge University Press, 2017.

[52] E. Tognini-Bonelli, *Corpus linguistics at Work*. Amsterdam: John Benjamins, 2001.

[53] R. Moon, *Fixed expressions and idioms in English*. Oxford: Oxford University Press, 1998.

[54] J. Yamashita**,** and N. Jiang, “L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations,” *TESOL Quarterly*, vol. 44, no. 4, pp. 647-668, 2010. DOI: https://doi.org/10.5054/tq.2010.235998.

[55] N. Nesselhauf, *Collocations in a learner corpus*. Amsterdam: John Benjamins, 2005.

[56] T. Johns, “From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning,” In T. Odlin (Ed.), *Perspectives on Pedagogical Grammar (Cambridge Applied Linguistics*, pp. 293-313. Cambridge: Cambridge University Press, 1994.

[57] P. Pérez-Paredes, and G. Mark, *Beyond concordance lines: Corpora in language education.* John Benjamins Publishing Company, 2021.

[58] D. Crystal, *Language and the Internet* (2nd ed.). Cambridge: Cambridge University Press, 2006.

AUTHORS' PROFILE

Amir H.Y. Salama is currently Associate Professor of linguistics and English language in the Department of English, College of Social Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Saudi Arabia. Also, he is a standing Professor of linguistics and English language in the Faculty of Al-Alsun (Languages), Kafr El-Sheikh University, Egypt. In 2011, Prof. Salama got his PhD in linguistics from the Department of Linguistics and English language at Lancaster University, UK. Since then, he has published at international journals like Discourse & Society, Critical Discourse Studies, Pragmatics and Society, Cogent Arts and Humanities, Semiotica, Corpora, and Translation Spaces. His research interests are systemic functional grammar, corpus linguistics, discourse analysis, pragmatics, cognitive semantics, translation studies, and semiotics.

ORCID: https://orcid.org/0000-0001-9320-558X

Waheed M. A. Altohami is currently Assistant Professor of English Language and Linguistics in the Department of English, College of Science & Humanities, Prince Sattam Bin Abdulaziz University (KSA). Also, he is a standing lecturer of English Language and Linguistics in the Department of Foreign Languages, Faculty of Education, Mansoura University, Egypt. His research interests include discourse analysis, cognitive semantics, corpus linguistics, and translation.

ORCID: https://orcid.org/0000-0001-8742-1366