Enhancing Sentence-Level Idiomaticity and Collocational Competence among Post-High school Saudi EFL Learners through the online Corpus of Contemporary American English

Awad H. Alshehri *

Department of English Language and Literature, College of Languages and Translation, IMSIU

Abstract

The use of Corpus of Contemporary American English (COCA) can enhance sentence-level idiomaticity of Saudi learners of English as a foreign language. It can also help develop learners' collocational competence by giving them access to the most frequent lexical bundles in a written context. Idiomaticity appears to be a key component for language learners particularly those who are expected to be involved in sentence writing tasks. To decide to what extent COCA can contribute to increasing EFLs' idiomaticity competence and enhance their writing skill, the researcher conducted some task-based activities involving writing at the sentence level. Learners were first tested to give feedback on a written text with some words and expressions miscollocated and misused (1A). Then, they were introduced to this online corpus analysis as a tool to explore and enhance idiomaticity and were given a list of words to explore using the corpus. Afterwards, in order to find out if the corpus could enhance their idiomaticity and collocational competence, they were given a similar test for analysis of word lexical bundles that are best used to covey meaning precisely as a native speaker would (1B). Finally, and in order to find out if the corpus could improve their idiomaticity and collocational competence in a new context, they were given a third test with expressions that are non-native, native and expressions that are acceptable in both languages. The findings show that COCA could increase language learners' awareness of the appropriateness or inappropriateness of collocational expressions. This could apply to other corpora such as iWeb and Glowbe.

Keywords: COCA, Collocational Competence, idiomaticity, Implicit Instruction, lexical bundles

INTRODUCTION

Idiomaticity and collocational competence have been considered a significant component for language learning and have accordingly received much attention by many scholars (Zahedi & Mirzadeh (2010). Idioms and collocations, being multi-word items, are essential elements for fluent, easy and streamlined production of language (Lewis, 2000 & Hyland, 2008). Such fluent production needs not only grammar but also a lexical item. Therefore, these two aspects appear to be inseparable as explained by Bahns (1993). This

* **Correspondence:** Awad H. Alshehri, Email: ahjsh@hotmail.com

view is supported by what Lewis (2008a) called lexicalized grammar as compared to grammaticalized lexis, the former being of much importance.

A coherent and cohesive text is usually made of meaningful and related chunks, as described by Lewis (2008b), which contribute to the overall felicity and appropriacy of text production. This means that a person's storage of these chunks is what forms raw data which facilitate perception of linguistic patterns, language forms, morphological structure and other linguistic features. What accounts for this is that language is stored in the speaker's mind as lexical bundles or chunks instead of single isolated words (Nattinger & DeCarrico, 1992).

There have been many studies on the topics of grammar, lexis and collocation, but only some of them tackled the relationship between writing idiomaticity and collocational competence on the one hand, and the use of online COCA on the other. This study, which is conducted in the Saudi context, attempts to find out whether the use of online COCA can enhance language learner's idiomaticity and collocational competence, filling what can be called a gap in the literature. Thus, the present study used short-term, task-based activities with the help of online COCA facility to find out if it can improve learner's idiomaticity and collocational competence and the effect of that on writing.

Collocational Competence and Idiomaticity

Profound knowledge of idioms and collocations and mastery of their usage constitute the key element of a native speaker's linguistics competence, which distinguishes a language learner from a native speaker. Perception and production of idioms and using collocations appropriately is not an easy task for a language learner due to the vast number of idiomatic structures and collocations, which urges the implementation and exploitation of the available digital systems that provide a rich environment for idiomatic structure and collocation learning. Digital systems make it easier to identify collocations in real documents available on the web which have been written or said by native speakers naturally in practical situations. The use of such techniques helps enhance the use and knowledge of idioms and collocations and forms the basis of language exercises if done under the teacher's control or even on the learner's own.

Considerable attention has been placed on the idea of collocations and idiomaticity being part of the discussions the phenomenon of formulaic language, particularly lexical bundles, as mentioned by Schmitt (2004) and Wray (2002). Therefore, it can be stipulated that an insufficient reservoir of lexical bundles could result in a language learner who is unable to select appropriate expressions whether in writing or speech (Kjellmer, 1990). In other words, a language learner is in need of storage of linguistic combinations that tend to appear frequently in written and spoken texts. These combinations are made of words and expressions that, statistically speaking, tend to co-occur in corpora, according to Durrant and Aydinli (2011). The interest in investigating frequently-occurring word combinations has witnessed an increase in linguistics for several decades, particularly after the advent of computer technology which, in turn, helped in the appearance of corpus-based methodologies. This change has made it much easier to investigative such linguistic phenomena, with a focus on language learning. The

use of online corpus helps the users to explore actual patterns of language and easily identify, at the lexical level, high-frequency collocations such as 'a high probability, do a lot of damage' and, at the syntactic level, colligations such as prepositional phrase 'in asset prices, for supervisory staff,'.

Collocations and idioms, therefore, can make efficient raw materials for classroom language instruction, where a language instructor can decide on what lexical bundles to teach by consulting corpora for a wordlist to find the lexical bundles that have a high frequency (Schmitt, 2012). These lexical bundles have come to light through corpus linguistics, as postulated by Schmitt, (2012). Biber and Birbieri (2007) describe lexical bundles as the "important building blocks of discourse" (p. 270) that have specific functions language-wise. These building blocks form the basis of idiomaticity and collocational competence, which, in turn, form the basis of a native speaker's formulaic language as explained by Schmitt and Carter (2004). For a non-native speaker, achieving idiomaticity and having adequate knowledge of conventionalized combinations is by no means a simple task.

Idiomaticity and collocational competence, being lexical bundles, contribute largely to the what scholars call discourse community. In this regard, Adel and Erman (2012) say that lexical bundles help in illustrating membership in the discourse community. This idea is also supported by the view that success in discourse community is attained by many interconnected and indispensable factors, such as the use of lexical bundles, which means that grammatical structures are not enough to produce acceptable speech, as described by Richards and Schmidt (2010). He believed that Idiomaticity is not merely speech that is grammatical; it is speech that is native-like in use, too. He gave an example of that as *"it pleases me that Harry was able to be brought by you"* which is grammatical but not native-like, while its counter-part sentence *"I'm glad you could bring Harry"* is grammatical and native-like and idiomatic, too (Schmidt, 2010, p. 270)*.

The ideas above can be practically expressed orally and in writing, being both parts of discourse. Both are important media of communication, which requires idiomaticity and collocation, making it a challenge for language learners to produce idiomatic discourse. Such discourse is what Adel and Erman (2012) describe as fluent and pragmatically correct in terms of language use. Such fluency and practicality can be examined and facilitated from the perspective of lexical bundles, which are the basic linguistic elements that serve significant functions in discourse construction (Biber, Conrad, & Cortes, 2004). This can be executed quite easily thanks to the significant advancements in technology which facilitates and provides statistical analysis of formulaic language.

One of these facilities, as explained by Callies (2013), is the COCA which has been used in language teaching, particularly EFL classroom, to enhance learners' understanding of how language pragmatically works. This online facility, in addition to other facilities, can improve students writing skills collocation-wise as alluded to by Nurmukamedov and Olinger (2013).

Saudi EFLs, as far as writing is concerned, proved to lack this idiomaticity and collocational competence, which appeared in some standardized test such as TOEFL and

IELTS, which urges further investigation. To the researcher's best knowledge, this topic has not been thoroughly researched in the Saudi context with the help of online facility such as COCA, which makes this topic worthy of research.

Vocabulary Teaching and Learning: Historical Development

Vocabulary teaching and learning has always been important in language pedagogy despite the fact that vocabulary-related research appeared to be insufficient (Meara, 1984). Some scholars, such as Wilkins (1974), indicated that vocabulary learning is as important as grammar itself. This led to some investigations of models of vocabulary knowledge such as Firth's (1957) modes of meaning. This type appears in the dictionary. Another mode is meaning by collocations, as pointed out too by Firth (1957). This means that the word meaning can be determined by its linguistic co-text more than what can be elicited from what is found in the dictionary. Firth's models paved the way for the development of other models that came afterwards. Another model called 'vocabulary knowledge model' was introduced by Lado (1957) in the book of *Linguistics across Cultures*. In this book, he emphasized the importance of vocabulary and provided some lexical elements such as form, meaning, frequency, and register, and he explained the difference between receptive and productive vocabulary knowledge.

In his book *Teaching and Learning Vocabulary*, Nation (1990) introduced another model that emphasized the productive and receptive knowledge of words. He explained that receptive knowledge of vocabulary is needed in reading while productive knowledge is used in writing. At the same time, he did not neglect the native speaker's knowledge of vocabulary. Nation mixed in his argument about vocabulary between knowledge of vocabulary and teaching, but his work was still significant in the field of language learning. He enlarged on his model by dividing word knowledge into form, meaning, and use.

Knowledge of vocabulary, therefore, is a significant part of the formulaic language, which was of supreme importance to scholars as we have seen so far. Therefore, language teachers, educators and researcher should give the formulaic language its proper weight, particularly those aspects related to idiomaticity and collocations being part of lexical bundles as described by Biber and Conrad (2004).

How COCA is used in this Study

Before explaining the use of COCA, the researcher thought it would make more sense for the students to examine a text where collocations and idiomaticity were deliberately written by the researcher with some mistakes or non-native like way. Then the students were familiarized with COCA interface, how to search words using online corpus and how to examine the frequency of their collocations in a variety of genres in spoken and written context by native speakers. They were expected to re-examine the texts given to them with the corpus where the words are used assumingly by native speakers. They were also expected to examine new texts with some collocational issue.

The researcher recommended the following stages for the students.

- Examine the texts provided to them with focus on collocations and idioms in terms of appropriateness.

- Making judgements about the appropriateness of the collocations and idioms used in the texts and suggesting better ways if any.

- Highlighting any idiomaticity and collocational issues in the texts, checking COCA for all possible genres in which the highlighted words occurred and choosing the most relevant lexical bundles that suit the text in hand.

- Making whatever corrections are needed to the text in hand.

- The researcher was expected to check the work and give feedback on improvement.

Philosophical Underpinning

There are various instructional approaches which serve specific purposes. The inductive approach is known to be student-centered, as opposed to the deductive approach. It makes heavy use of the strategy described as noticing strategy so instead of the teacher explaining a certain concept or idea and providing examples afterwards, the teacher starts with supplementing many examples that show how the concept or idea are used. The students are expected to use the strategy of noticing to using the examples to find out how the concept or idea is used. For example, if the teacher intends to teach a grammatical rule of some type, s/he will provide the students with a variety of examples where the rule is used. They are expected to find out, by noticing, how the rule is implemented and then determine what the grammar rule is. This activity can be concluded by an explanation of the rule by the students to make sure they have achieved the target, and it can be followed by feedback from the teacher if need be.

Noticing is an indirect awareness raising which is considered implicit instruction as explained by Ellis (2015). Noticing is an element of comprehension can take place through modified input and corrective feedback through the exposition to real-life examples written by native speakers. What happens, in fact, is an interactional modified input as described by Long (1981). When learners examine a text, with a focus on collocations for example, and then they notice how native speakers collocate these words, they undergo what we can call input modification. This input-based task is based on the noticing of certain linguistic patterns which the teacher intends for the students to acquire. This type of instruction can assist learners of different levels of proficiency to acquire and develop their linguistic competence.

As regards Saudi students, they, like any other EFLs, must have some deficiencies in the use of idioms and show some lack in collocational competence as shown by the analysis of their writing such as *form myself*, *heavy tea*, *complete answer*, *expensive advice* just to name a few. Unsurprisingly, Arabic-speaking EFLs have collocation errors due to the idiosyncrasies and the vast scope of English lexical items and collocations, as described by Thomas (1984).

Purpose of the Study

This study aims to examine the potential of using online facilities, namely COCA, to enhance students' writing through the use of appropriate collocations and idioms as used by native speakers, which was, at this stage, at the sentence level. This will hopefully simplify and promote their performance in two ways: autonomy in selecting appropriate collocations, and self-assessment of their writing. Their use of COCA will also give indications about the efficiency of the app and their ability to use this online facility, and probably other similar facilities, to promote their autonomous interpretation of online corpora which are relevant to their needs.

METHOD

The researcher designed and applied three online test-like instruments, 1A, 1B, and 2, each of which included 15 sentences having certain collocations, some of which were formed correctly in native English while some had some L1 idioms and collocations which did not fit the English text. 1A and 1B included the same items to see if there was any improvement in the students' performance after using COCA. The students were requested to do the first test (1A) and choose "appropriate", "inappropriate" or "I am not sure" and then submit their replies. Later on, they were introduced to the COCA and were requested to redo the same test while consulting the COCA. After that, they were requested to use the COCA and do two more different tests (1B) which had the same items, and (2) which had new items. The aim of the 1A was to find out advancement in their performance, and the aim of (2) was to see if they students could apply COCA to new idiomatic expressions. They were given these tests to find out the extent to which the student could benefit from COCA to enhance their idiomaticity and collocations. The results were analyzed using Excel facility. The researcher checked and evaluated their performance to answer the following question: To what extent can EFL students benefit from COCA?

Study Sample

The researcher chose to apply this study to only a few post-high school students since it is more of an experimental study. They were all post-high school students enrolling in university and were getting ready to major in English and are supposed to have acceptable writing skill. Their ages ranged between 19 and 20, and they have passed the entrance exams required for the college of languages and translation.

STUDY RESULTS

Numerical analysis of the first activity called 1A is shown in Table 1 below:

Table 1. Collocations and idiomaticity 1A

No	Item	Appropriate	Not sure	Inappropriate	Correct
1	She had a fast glance at the book	75%	0%	25%	25%
2	She had a fast meal in the	83%	0%	17%	17%
3	It was a complete eclipse	83%	17%	0%	0%
4	We have bread and salt among	83%	17%	0%	0%
5	Some people transport news	67%	0%	33%	33%
6	We had heavy tea	83%	17%	0%	0%
7	He is alive and being provided	25%	25%	50%	50%
8	I made a walk to the market	83%	17%	0%	0%
	Average	73%	11%	16%	16%
9	I made a trip to the club	67%	8%	25%	67%
10	It was a total eclipse	58%	25%	17%	58%
11	He paid us a visit	58%	17%	25%	58%
12	He is a heavy smoker	33%	25%	42%	33%
13	We saw a school of whales	0%	33%	67%	0%
14	He is alive and kicking	42%	25%	33%	42%
15	He made a quick meal	92%	0%	8%	92%
	Average	**50%**	**19%**	**31%**	**50%**
	Overall Average		**33%**		

Table 1 shows two sets of phrases. 1-8 are phrases that are formed with L1 one in mind, meaning that L1 interference could have played a role and influenced the respondents' answers. Most of the respondents accepted these phrases despite the fact that native speakers or even people from different cultures would consider them awkward and may not understand the intended meaning. Most of the respondents accepted the expressions "fast glance", "fast meal", "a complete eclipse", "bread and salt", "transport news", and "make a walk" and they thought that these words would collocate appropriately. At the same time, all the respondents rejected the correct English collocation "a school of whales". Other negative scores ranged between 17% to 67% since most of them accepted the expressions "fast glance", and "fast meal". Only 42% accepted the collocation "alive and kicking", and only 50% rejected the wrong expression "alive and being provided for". The overall average of the positive responses was 33% in this test, and 15% were not sure or did not know.

Numerical analysis of the second activity test 1B is shown in Table 1 below:

Table 2. Collocations and idiomaticity 1B

No	Item	Appropriate	Not	Inappropriate	Correct
1	She had a fast glance at the	25%	17%	58%	58%
2	She had a fast meal in the	58%	17%	25%	25%
3	It was a complete eclipse	67%	17%	17%	17%
4	We have bread and salt	42%	25%	33%	33%
5	Some people transport	67%	0%	33%	33%
6	We had heavy tea	33%	17%	50%	50%
7	He is alive and being	17%	25%	58%	58%
8	I made a walk to the	33%	25%	42%	42%
	Average	43%	18%	40%	40%

9	I made a trip to the club	75%	8%	17%	75%
10	It was a total eclipse	67%	25%	8%	67%
11	He paid us a visit	75%	17%	8%	75%
12	He is a heavy smoker	42%	25%	33%	42%
13	We saw a school of whales	50%	17%	33%	50%
14	He is alive and kicking	58%	25%	17%	58%
15	He made a quick meal	100%	0%	0%	100%
	Average	67%	17%	17%	67%
	Overall Average			**53%**	

Table 2 also shows the same two sets of phrases mentioned in Table 1 above. 1-8 would collocate correctly in Arabic, while 9-15 are collect English collocations. The average correct responses to the first set of phrases improved by 24% from 16% to 40% after consulting COCA. The average correct responses to the second set improved by 17% from 50% to 67%. Consequently, the overall average of the positive responses improved by 20% from 33% 53% in this test and 17.5% were still not sure or did not know.

Table 3. Collocations and idiomaticity 2

No	Item	Appropriate	Not	Inappropriate	Correct
1	She had a short look at the	25%	25%	50%	50%
2	She sometimes tries strange	58%	25%	17%	17%
3	We did a journey to the forest.	58%	25%	17%	17%
4	We did a run before	42%	33%	25%	25%
5	It is my job as a reporter to	42%	17%	42%	42%
Average		45%	25%	30%	30%
6	Don't put all your eggs in one	92%	8%	0%	92%
7	Last night's match is the press	83%	8%	8%	83%
8	We met a group of wolves.	100%	0%	0%	100%
9	It was a hit and run accident.	75%	8%	17%	75%
		88%	6%	6%	88%
10	She had a brief look at the	67%	17%	17%	67%
11	We made a journey to the	58%	33%	8%	58%
12	We went for a run before	67%	25%	8%	67%
13	She sometimes tries exotic	58%	25%	17%	58%
14	It is my job as a reporter to	42%	25%	33%	42%
15	We saw a pack of wolves.	50%	25%	25%	50%
	Average	57%	25%	18%	57%
	Overall Average			**43%**	

Table 3 shows three sets of phrases that are entirely different from the phrases in 1A and 1B. 1-5 are phrases that may sound acceptable to Arabic-speaking language learners, while 10-15 are collocated correctly in English despite the fact that some learners may consider weird. The set 6 to 9 includes phrases that are acceptable in Arabic and English. The average score in the first set was higher than that of 1A and lower than the score in 1B. Moreover, the overall average of the positive responses was 43%, which is also higher than that of 1A and 1B.

FINDINGS

The results of 1A and1B, which had the same items, yielded the following findings. The utilization of COCA raised students' awareness of the inappropriateness of some L1-based expressions, increasing from 16% to 40%. It also helped in familiarizing them with new English collocations and expressions in which their performance increased by 17% from 50% to 67%. The results of test 2, which had entirely new items, also produced positive results. Awareness of the inappropriateness of L1-based expressions increased by 14% from 16% to 30%, and their awareness of the appropriateness of the English collocations increased by 7% from 50% to 57% since the items of the second test were new. However, their performance in expressions and collocations that are accepted in both languages scored 88%, which logically was supposed to be 100%. This could be attributed to hesitation and in confidence.

CONCLUSION

We can infer from the findings that this sample had not been exposed systematically to the idioms and collocation of the target language, which might be considered as a general phenomenon for Saudi students at the level of pre-college education. The introduction of COCA did proof to have a positive impact on improving students' awareness of the fact their target language has its own idioms and collocations that can be learned by looking at and benefiting from electronic corpora such as COCA, iWeb, and Glowbe. These online facilities can simplify language learning and master the pragmatic use of its collocations, as explained earlier by Callies (2013). Collocations, as explained by Nurmukamedov and Olinger (2013), should be practiced in EFL classrooms to improve students' writing skills, too. Deep knowledge of L2 idioms and its collocations does constitute the key element of a speaker's linguistics competence which helps a language learner to easily identify collocations in real printed or online documents written or said by native speakers in a natural setting. Collocations, as mentioned by Schmitt (2004) and Wray (2002), are a key part of lexical bundles. Learners' inadequacy in the lexical bundles reservoir could result, as believed by Kjellmer (1990), in their inability to select appropriate expressions.

Therefore, collocations and idioms should be part of the raw material used in any language classroom, where students can consult corpora to find the lexical bundles that have high frequency, as suggested by Schmitt (2012). Mastering collocations and idioms can lead to acceptable and successful integration with the community that uses conventional lexical bundles, as described by Richards and Schmidt (2010), since the linguistics co-text of a given word determines its meaning more precisely than what a dictionary does (Firth, 1957).

REFERENCES

Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non- native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81- 92.

Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, 47(1), 56-63.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics,* 25(3), 371-405.

Callies, M. (2013). Advancing the research agenda of interlanguage pragmatics: The role of learner corpora. In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2013: New Domains and Methodologies* (pp. 9-36). New York: Springer.

Durrant, P., & Aydinli, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for specific Purposes, 30, 58- 72.*

Ellis, R., (2015). *Understanding Second Language Acquisition, -Oxford Applied Linguistics*. Oxford University Press.

Firth, J. R. (1957). *Papers in linguistics, 1934- 51.* London: Oxford University Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27(1), 4-21.*

Kjellmer, G., (1990). A mint of phrases. In K. Aijmer, & B. Altenberg (Eds.), *English corpus linguistics: Studies in honour of Jan Svartik* (pp.111-127). London: Longman.

Lewis, M, (2000). *Teaching collocation: Further developments in the lexical approach.* England: Language Teaching Publication.

Lewis, M. (2008a). *The Lexical approach: putting theory into practice.* England: language Teaching Publication.

Lewis, M. (2008b). *Implementing the lexical approach: Putting theory into practice*. England: Language Teaching Publication.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers.* Ann Arbor: University of Michigan Press.

Long, M. H. (1981). Input, interaction, and second-language acquisition. *Annals of the New York academy of sciences*, *379*(1), 259-278.

Thomas, H. (1984). Developing the stylistic and lexical awareness of advanced students. *English Language Teaching Journal, 35*, 187-191.

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* New York: Newbury: House Publishers.

Nattinger, J., & DeCarrico, J. (1992). *Lexical Phrases and Language Teaching.* Oxford: OxfordUniversity Press.

Nurmukhamedov, U., & Olinger, A. R. (2013). Computer-mediated collocation: Resources for exploring word choice in English academic writing. *Writing & Pedagogy,* 5(1), 121-150.

Orenha-Ottaiano, A. (2012). English collocations extracted from a corpus of university learners and its contribution to a language teaching pedagogy. *Acta Scientiarum language and culture*, 34(2), 241-251.

Richards, J.C. & Schmidt, R. (2010). *Longman dictionary of language and applied linguistics.* London: Longman.

Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing and use*. Amesterdam: John Benjamins.

Wray, A. (2002). *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press. Zahedi, H., & Mirzadeh, M. (2010). Collocational cloze. The effect of deletion procedure and gender. *Iranian EFL Journal*, 6(1), 143-157.

Zarei, A. (2002). *Patterns of Iranian advanced learner's problems with English collocations.* Unpublished doctoral dissertation, University of Isfahan, Iran.