

The Use of COCA to Promote Writing Idiomaticity among Advanced Iranian EFL Learners

Aylar Fallah Vazirabad, Dr. Awad Hajran Alshehri

Department of Foreign Languages and Literature, University of Tabriz, Tabriz, Iran
School of Modern Languages & Applied Linguistics, University of Limerick, Ireland
Fallahaylar@Tabrizu.ac.ir, Awad.alshehri@ul.ie

Abstract: *The Corpus of Contemporary American English (COCA) has never been examined in an Iranian English as a Foreign language (EFL) classroom to develop students writing skills. However, it might allow learners to explore collocational competence by giving them access to the most frequent lexical bundles in written context. Mastery of idiomaticity of language is pivotal in the training of advanced level students who are preparing for IELTS writing tasks. To determine how far the COCA can contribute to increase this awareness, a series of task-based activities involving writing was drawn up and carried out in an EFL class. They were first introduced to this corpus analysis tool and encouraged to explore it further. Later on, in order to complete a writing task, they were prompted to resort to a series of strategies to collect information, to analyse and interpret data, and to draw conclusions about the most appropriate 2- 5 word lexical bundles of language interaction that can be used in their relevant setting. This paper provides (a) a brief literature review on this topic (b) a description of task-based activities, and implementation process, the students' strategies, the evaluation and self-reflection procedures, and (c) a critical reflection on this study that may open the path for further developments in this area and (d) an explanation of Implicit Instruction and the significant desire for future research. Finally, the questionnaire elicits students' opinions about the new tools' implementation, and the outcomes are discussed in detail.*

Keywords: *Corpus analysis tools, idiomaticity, lexical bundles, written discourse, advanced EFL learners, Implicit Instruction*

1. INTRODUCTION

According to Zahedi & Mirzadeh (2010) there is significant attention to second language learners' collocations and idioms. Lewis (2000) emphasized the importance of multi-word items compared to single words. Hyland (2008) highlights that multi-words illustrate components of fluent production and an important aspect of language learning. Bahns (1993) noted that grammar and lexical items are inseparable. Lewis (2008a) states the significance of lexicalized grammar compared to grammaticalized lexis.

According to Lewis (2008b), language consists of meaningful chunks that is required in producing coherent text. Repertoire of chunks which forms raw data by which learners start to perceive patterns, forms, morphology and other linguistic features. Nattinger and DeCarrico (1992) explains the language is stored as lexical bundles (chunks) rather than single words, sentence heads and institutionalized sentences.

Even though several studies have been done in the area of grammar, lexis and collocation, very few of them have investigated the relationship between teaching lexical collocation and EFL learners' writing proficiency (Ghonsooli et al., 2008; Mounya, 2010). This study tries to fill this gap. Thus the present study investigated the effect of digital lexical collocation tool (i.e., noticing, highlighting, and consciousness-raising) on learning lexical collocations and the impact of lexical collocation instruction on developing writing proficiency of advanced EFL learners.

2. THE IMPORTANCE OF IDIOMATICITY

The phenomenon of formulaic language in general and lexical bundles in particular, has received great deal of attention (Schmitt, 2004; Wray, 2002). According to Kjellmer (1990), individuals with an insufficient insider experience may not be able to use the most appropriate expression, and hence their speech may not seem completely adequate.

According to Durrant and Aydinli (2011), in frequency based approaches, those linguistic combinations which have a tendency to appear with high frequency in texts, are defined as formulae that are strings of linguistic items (including words, parts of speech, and semantic fields), which have a statistical tendency to co-occur in corpora. Examples include high-frequency collocations (e.g., a high probability; do a lot of damage); colligations (e.g., prepositional phrase: for supervisory staff; in asset prices); semantic preferences (e.g., cause & effect; cause operating risks); and semantic prosody (e.g., caused happiness).

By using corpus linguistics, researchers can explore actual patterns of language use and also develop materials for classroom language instruction (Schmitt, 2012). Similarly, a teacher can decide what sequences to teach by consulting a wordlist from corpora by finding the frequencies of lexical bundles. Reppen and Simpson-Vlach (Schmitt, 2012, p. 98) indicate that “Lexical phrases, or lexical bundles, are an area of collocation studies that has come to light through corpus linguistics.

Lexical bundles are considered as “important building blocks of discourse” (Biber & Birbieri, 2007, p. 270) that serve specific functions. The knowledge of conventionalized combinations of words is essential to achieve idiomaticity. This knowledge comes to the mind of the native speakers automatically, and it has come to be known by the term ‘*formulaic language*’ (Schmitt & Carter, 2004). Achieving idiomaticity, or in other words, the knowledge of conventionalized commixture of words is a challenging task. The other important contribution of the knowledge of lexical bundles is illustrating membership in a discourse community (Adel & Erman, 2012). Moreover, there is no attention paid to lexical bundles through corpus analysis tools in Iran. Therefore, the outcomes of this research might be beneficial in this regard.

Success in discourse community is achieved through many interrelated and indispensable factors. One of those factors is to use lexical bundles in discourse and achieve idiomatic discourse both in writing. Richards and Schmidt (2010, p. 270) put it:

Idiomaticity is the degree to which speech is not simply grammatical but also native-like in use. For example, “it pleases me that Harry was able to be brought by you” is grammatical but not native-like or idiomatic, whereas “I’m glad you could bring Harry” is both grammatical and idiomatic.

Writing, as parts of discourse, is an important media of communication, and finding optimal ways of making writing idiomatic is a challenging task for learners. Idiomatic discourse has features such as fluency and pragmatically correct language use (Adel & Erman, 2012). Nowadays, it is possible to examine idiomaticity from the perspective of lexical bundles. Lexical bundles are basic linguistic constructs that have important functions in the construction of discourse (Biber, Conrad, & Cortes, 2004). One of the important advances in analysing formulaic language is using statistical information through utilization of concordance programs. The line on lexical bundles research makes use of this technological advancement led by Biber, Johansson, Leech, Conrad, & Finegan, (1999).

The COCA (<http://www.americancorpus.org>) has been used in the EFL classroom to help learners better understand how language works at different levels of analysis (Bennett, 2010; Boulton, 2011; Callies, 2013; Orenha-Ottaino, 2012). It has also been used to enhance their text production and develop their writing skills (Kim, 2009; Nurmukamedov & Olinger, 2013). This is linked to idiomaticity and required skills needed for IELTS test of Iranian through digital era which lacks familiarity and further investigation.

To the best of the researcher’s knowledge, computer- retrieved idiomaticity in learners’ discourse have been under-investigated by applied linguists. Therefore, the lack of research in this context makes the upcoming research worth noticing.

3. PRIMARY AND LATER MODELS OF VOCABULARY KNOWLEDGE

The first model of vocabulary knowledge belongs to Firth’s (1957) modes of meaning found in dictionaries. To quote Firth (1957): ‘the lexical meaning of any given word is understood by different statements of meaning as various levels’ (p. 192). For example, the typological or grammatical levels which are found in dictionary are related to those levels of meaning. Firth (1957) pointed out to another model, that is, meaning by collocation. He stated that the meaning of a word was determined by its linguistic co-text, and not so much by the meaning which was found in dictionary. These

models acted out as the building blocks of the development of other models for future. For example, later models such as Richards (1976) and Nation (1990) were the expansion of Firth's (1957) previous models. At the same time, Lado (1957) introduced his vocabulary knowledge model in his famous book of *Linguistics across Cultures*. Like other researchers, he mentioned that vocabulary needed more attention, and provided teachers with some lexical elements that he thought were essential for teachers. He touched on many form, meaning, frequency, and register in his great work, and on the other part of his model he distinguished between receptive and productive vocabulary knowledge. Although it was a comprehensive study, it was not without shortcomings. For example, his model's building block was the contrasting of two vocabulary systems of two different languages, rather than being based on the complete components of vocabulary knowledge.

Nation (1990) presented another pioneering model in his book '*Teaching and Learning Vocabulary*'. The model pointed out to issues of receptive and productive knowledge of words; in other words, whether a learner needed to use the vocabulary in reading, which is then receptive, or in writing which is productive. Nation (1990) paid attention to the needs of the students, but at the same time based his model on the knowledge of native speakers' knowledge of vocabulary. A limitation of his study was his mixing of vocabulary knowledge and teaching indication. In spite of the limitations, his work was so effective on the field that most studies mentioned it on research. On his later model which was the development of the first one, he reorganized the work into form, meaning, and use. When it came to meaning, the association of the word with specific meaning, the concept of the words and their synonyms were discussed. In regard to use, the grammatical functions of the words such the patterns they occurred were described. Also, he discussed the constraints such as when, where, and how the words should be used. It was a clear fact that the importance of formulaic language was of utmost importance to researchers as it was seen in the reviews. As a result, the formulaic language should find its appropriate position in linguistic analyses and in fields which relate to language teaching. As it is evident, there is a great emphasis on collocations and formulaic language in the teaching of vocabulary, which connects those reviews to the studies of formulaic language. Moreover, collocations are different from lexical bundles in terms of the number of words, because lexical bundles consist of more words in relation to collocations (Biber & Conrad, *ibid*). I am going to review corpora studies, which helps researchers to find the different kinds of formulaic language in different registers. As this study also uses corpus methodology, I also elaborated it on the following sections.

4. FUNCTIONS AND BASICS OF COCA

The teacher explained how to search the electronic database of written texts so that students can identify and examine frequency of certain words (strings of words). They will be familiarized with how to search for collocates, words, lists of parts of speech and synonyms (synonym collocates lists) or to compare inquiries by searching variety of genres in spoken and written context. Teacher will also examine their understanding by checking their writing with the corpus at hand. They will realize that it is often dangerous to look at only the frequency count to decide whether to use the collocation. Therefore, having similar frequency counts does not mean both words are possible in a given context, and this leads to interpretation.

Taxonomy of stages and modes recommended was as follows:

Stage 1: Finding out and highlighting how required words collocates in a particular type of setting and its frequency, and self-assessment of students' writings based on the outcomes.

Stage 2: Comparing and contrasting variety of displays of collocates and the best version that fits the required writing task in hand based on the format (e.g. academic or spoken)

Stage 3: Following and collecting as much information as possible about the genre in which they occurred and selecting the most relevant lexical bundles for the text in hand. This will teach them to select the best lexical bundles from one authentic context and use it in their own context without making any plagiarism, and students were also asked to highlight the change to their own writing based on COCA. This was illustrating how much COCA was affective in promoting their experience and in which ways.

Stage 4: Frequency of data analysis to determine how influential/significant was the bundles throughout the upcoming years, ‘wildcards’ and how best to use them in new setting.

Stage 5: How best to familiarize them with technology and how to search the required items.

The above strategies and basic instructions were given to the students in the first few sessions of the course and as worksheets. The above strategies had to be applied in the completion of the following tasks:

- Writing IELTS Task 1 and 2
- Preparing for an Interview Session

5. THE ROLE OF IMPLICIT INSTRUCTION

According to Rod Ellis (2015) implicit instruction takes place throughout ‘noticing’ and indirect awareness raising. Comprehension happens through pre - modified input, corrective feedback and pushed output by the learner. Input-based tasks are based on comprehension and noticing of linguistic patterns which needs to be acquired. Brainstorming and pre-tasks in this regards are strategies in which the facilitator only exposes the learner to the pre-modified tasks from which learners comprehend the new patterns and linked this to their resource-dispersing variables of the task. This task-based instruction assists learners specially learners with advanced level of proficiency.

There is growing evidence that form-meaning connections can be learned implicitly – vital support for the argument that learning such a complex system is possible without deduced knowledge of rules. However, according to Mitchell et al (2013) much more research is needed to determine exactly what can be learned implicitly, by whom and when. Determining the type and amount of input that is necessary for implicit learning to occur presents considerable challenges for future and aimed for current research.

6. STATEMENT OF THE PROBLEM

As far as difficulties in idiomaticity and collocational use are concerned, Iranian learners of English are by no means an exception. Any analysis of students’ speech or writing indicates a deficiency in this regard. Despite having sufficient lexical or grammatical knowledge, most Iranian EFL learners seem to experience serious problems with the production of collocational patterns. Such erroneous expressions as *strong rain*, *to take birthday*, *heavy tea*, *to begin a family*, *hard question*, just to name a few, are not due to poor mastery of grammar or lexis. These problems, as Koosha and Jafarpour (2006) have pointed out, stem largely from a lack of appropriate collocational knowledge among Iranian EFL learners, the inadequate emphasize given to collocational patterns in their textbooks, and the type of instructions they receive. Moreover, such multi-word expressions have not usually been the focus of teaching. The fact is that research professionals in our country and most of the research carried out to date (Akbari, 1995; Zarei, 2002; Koosha & Jafarpour, 2006) have investigated the patterns and usage of collocation in the learners’ written product, and only a few attempted to solve the problem.

7. PURPOSE OF THE STUDY

This study aims to familiarize advanced level Iranian students with COCA tool to promote their autonomy in self-assessment and writing skills (Collocations, colligation, and grammaticality) during the preparation course for IELTS exam. The goal of the experiment is:

- To examine the effect of COCA in self-correction and autonomy in preparation for writing tasks required for IELTS.
- To examine COCA’s friendliness, appropriateness and functionality.
- To improve their autonomous interpretation of the corpus data relevant to the needs.

8. PARTICIPANTS

Twenty students participated in this study. They were all students from two IELTS classes (Writing) at Namavaran Language Institute in Tabriz, Iran. The participants in this study included adult students

who were studying English to improve their writing skills to participate in IELTS exam in order to continue their further education abroad. They were between 22-38 years of age. Overall, they had an IELTS score of 6.5, and they were expecting to gain a score between 7-9.

9. METHODOLOGY

We utilized a questionnaire consisting of 10 questions to explain how they were able to use the tool for their advances. The questionnaire was interested to answer the following topics: impact of the tool in the writing process, effect of the tool in providing the required lexical collocations for the idiomaticity, grammaticality and adopting native-like competence from the authentic corpora to the new setting, and usefulness of the new tool to improve their autonomy and self-evaluation of their own written task in hand in a game- like setting and in their convenience.

10. RESEARCH OUTCOMES

In the texts produced by the students, significant progression was detected overall to be random idiomatical expressions or at the level of prepositional phrases, relative clauses, article, tense, active, passive voice and adjusting grammaticality and bundle typologies such as Noun phrase with of clause fragment, noun phrase expressions and noun phrase with other post-modifier fragments respectively, specially in writing task 2. Therefore, students produced frequently new collocations that fit their new writing context and created a more advanced level text and avoided grammatical errors, and overall comparing the first writing and final tasks at hand. The students adopted the new lexical bundles in their writing style whether it was simple or complex in general despite some inauthentic choices of vocabulary in the new setting which needs further progress in general most of the choices were acceptable. The students progressed in finding synonyms for word list and therefore avoided repetition of words in their writings. They were able to use the ideas from COCA and adapt it to their main ideas and context at hand and to present coherence and relevancy throughout the text making the text native-like. The students who answered to the questionnaire illustrate overall positive view about potential of aid of COCA as their self-concordance tool in writing process. Nearly 80% of the students selected 'sure' for using the tool in further writing sessions and 20% were uncertain as they still used collocation paper dictionary or no comment at all. In relation to learning new collocates with COCA, the students ranked adopting the collocates from COCA to their new context as the most significant usage of the tool in their writing and shaping their arguments. 80% of the students ranked it as the first and most useful tool in writing. 60 % of students mentioned that friendliness, appropriateness and functionality specially in the process of giving and creating novice sentences, ideas and paraphrasing is what made the tool unique. With regard to the usefulness of the tool in providing relevant lexical bundles, 60% of students still believed that the tool could divide an individual genre for successful IELTS writing, especially statistics and arguments as the most important leading tool. All the twenty students expressed that they had no previous experience of the tool or being suggested by their instructor. Finally, 90% confirmed that the COCA promoted their writing skills for IELTS exam and teacher scoring.

11. CONCLUSION

In Iran, there is no attention to solving the problem of collocational competence. There are no textbooks designed for that, no teacher training and no tool in this regard, but the research results of the students' written and spoken texts show that there is an increasing number of collocational errors. In this research, therefore, the students got familiar with the new tool that gave them a clear view of the authentic collocations, lexical bundles, idiomaticity and synonyms from a real-life language. To use it efficiently, they were familiarized with the technics trough basic instructions and worksheets. They were asked to adapt their previous knowledge with the one at hand, and the tool offered substantial bundles as the reflection of the new possibilities in their writing. They had difficulty in coping with the technical aspect of the tool as well as the best ways of interpreting the old tool based context for their writing at hand. They were shown creative and new ways by the instructor in which their ideas could be written with creative and new idiomatic structures. They become familiar with evaluating and selection procedures. As far as the tasks were concerned, both their writing assessment and questionnaire show a positive impact of the COCA tool in providing additional ideas as well as idiomatic expressions which boost their paraphrasing and overall a positive affect in aiding their future writing progress.

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.
- Akbari, R. (1995). The role of language-specific collocations on the written production of intermediate Iranian students learning English as a foreign language. Unpublished MA thesis, Tarbiat Modarres University, Tehran, Iran.
- Bahns, J. (1993). Lexical collocations: A contrastive view. *ELT Journal*, 47(1), 56-63.
- Bennett, G. R. (2010). Using corpora in the language learning classroom: Corpus linguistics for teachers. Ann Arbor: University of Michigan. *New Trends in Computer-Assisted Language Learning: Working Together* (pp. 39-46). Madrid: Macmillan ELT.
- Biber, Johansson, Leech, Conrad, & Finegan, (1999). *Longman grammar of spoken and written English*. Harlow; Pearson Education.
- Biber, D., & Barbbieri, F. (2007). Lexical bundles in university and spoken registers. *English for Specific Purposes*, 26(3), 263-286.
- Biber, D., Conrad, S., & Cortes, V. (2004). 'If you look at...': Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405.
- Boulton, A. (2011). Language awareness and medium-term benefits of corpus consultation. In A. Gimeno Sanz (Ed.), *New Trends in Computer-Assisted language learning: Working Together* (pp. 39-46). Madrid: Macmillan ELT.
- 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 (2011) 58- 72*. Available online at www.sciencedirect.com.
- Firth, J. R. (1957). *Papers in linguistics, 1934- 51*. London: Oxford University Press.
- Ghonsooli, B, Pishghadam, R, & Mahjoobi, F. (2008). The impact of collocational instruction on the writing skill of Iranian EFL learners: A case of production and process study. *Iranian EFL Journal*, 2, 36- 59.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Kim, Y. J. (2009). *Effectiveness of on-line corpus research in L2 writing: Investigation of proficiency in English writing through independent error correction*. Master thesis. University of North Texas. Retrieved from <http://digital.library.unt.edu/ark:/67531/metadc12140/m1/1/>.
- 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.
- Koosha, M., & Jafarpour, A. A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal*, 8(8), 1-14.
- 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.
- Lyons, Johns, 1967. Firth's theory of meaning. In memory of J.R. Firth, ed. By C. E. Bazell et al., 288-302. London: Longmans.
- Mitchell, R. Myles, F. and Marsden, E. (2013). *Second language Learning Theories*. (Third Edition). Routledge Taylor and Francis Group. London and New York.

- 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: Oxford University 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. Retrieved from <http://www.periodicos.uem.br/ojs/index.php/ActaSciLangCult/article/view/17130/pdf>.
- Richards, J.C. & Schmidt, R. (2010). *Longman dictionary of language and applied linguistics*. London: Longman.
- Schmitt, N. (2004). *Formulaic sequences: Acquisition, processing and use*. Amsterdam: 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.