

## Collocation Networks of Selected Words in Academic Writing: A Corpus-Based Study

**Eman Adil Jaafar**

Department of English  
College of Education for Women  
University of Baghdad, Iraq  
Email: [eman\\_jafer@coeduw.uobaghdad.edu.iq](mailto:eman_jafer@coeduw.uobaghdad.edu.iq)

Received: 1/4/2021

Accepted: 3/10/2022

Published: 3/24/2022

### Abstract

This study aims at shedding light on the linguistic significance of collocation networks in the academic writing context. Following Firth's principle "You shall know a word by the company it keeps." The study intends to examine three selected nodes (i.e. research, study, and paper) shared collocations in an academic context. This is achieved by using the corpus linguistic tool; GraphColl in #LancsBox software version 5 which was announced in June 2020 in analyzing selected nodes. The study focuses on academic writing of two corpora which were designed and collected especially to serve the purpose of the study. The corpora consist of a collection of abstracts extracted from two different academic journals that publish for writers from different countries around the world. This corpus-based study aims at examining the significance of chunks of language in texts. The concept of collocations is crucial in corpus linguistics to identify semantic relations. This can help in the teaching and learning processes. Furthermore, this study is conducted to answer the following research questions; first, whether the three words *study*, *paper*, and *research* are used interchangeably in the corpora or not? Second, what are the shared collocational associations surrounding the selected nodes? Finally, it is worth noting that the study of collocations highlights the linguistic features of texts through computational analytical tools that can save time and help to gain objective results systematically. The findings show that 'research' and 'study' are used rather interchangeably in the writing of the abstracts; however, 'paper' has fewer shared collocations in the same academic context.

**Keywords:** academic writing, collocational networks, corpus linguistic, GraphColl, linguistic significance, node

**Cite as:** Jaafar, E.A. (2022). Collocation Networks of Selected Words in Academic Writing: A Corpus-Based Study *Arab World English Journal*, 13 (1) 240-255.

DOI: <https://dx.doi.org/10.24093/awej/vol13no1.15>

## Introduction

Knowing collocation networks is an essential part of linguistic competence. Identifying words relations has an immense significance in language learning because it brings the learners' attention to words relations. However, Nagano and Kitao (2007) stress that the study of collocation, in general, is neglected or has been given less attention in comparison with other aspects of the English language such as grammar, phonetics, and phonology in English as a foreign language (EFL)/ English as a second language (ESL) context. Nagano and Kitao further illustrate that learners even advanced ones misuse collocations due to the influence of their first language. The researcher here agrees with these insightful points and argues that this difficulty of dealing with collocation is due to the limited exposure to authentic learning situations. Thus, this paper presents a study of collocations in their real authentic context. It is believed that learning any aspects of language deprived of their real situation will lead to unsatisfying results. From this point, it is hypothesized that learners can achieve their goal of learning collocation through GraphColl enables them to identify words in context and observe semantic relations in a systematic, objective way (Brezina, 2018). In this paper, the main aim is to show what are the associations related to the nodes; *research*, *study*, and *paper*? First, a historical background will be presented to reveal the significance of the study of collocation. Thus, the current study aims at answering these basic research questions: first, do the three words study, paper, and research are used interchangeably or synonymously? Second, what are the shared collocational associations surrounding the selected nodes? Thus, this article presents a practical study to identify and explain collocation networks within academic writing.

## Literature Review

### *Definitions and Historical Background*

Baker (2016) demonstrated that studies on collocation networks using GraphColl give detailed and rigorous analysis opposite to the traditional studies that focus on 'pairwise relationships' among words. Similarly, Gablasova, Brezina, and McEnery (2017) ascertained that collocational network as a unit of formulaic language provides useful insights about language learning not only in the first language (L1) but also in second language (L2) language production. Observing collocational relations or cross-associations of the words in certain text/s or discourse can describe the 'aboutness' of the examined data (Brezina, 2016).

Furthermore, El-Dakhs (2015) confirms that "collocations enormously contribute to efficient language comprehension and production. Regarding comprehension, memorizing chunks helps learners predict the content of texts, and automatically enhances comprehension". In this respect, Davis & Kryszewska (2012) confirm that readers can understand one meaning or two when a word stands alone or in isolation. However, we can get a broader significance when the same word is in company with other words.

Sinclair (1991, 2004) defines the concept of collocations as "the occurrence of two or more words within a short space of each other in a text" (p. 170) usually the span of four words to the left and right of a selected node. Furthermore, Nesselhauf (2005) explained that the term collocation refers to "some kind of syntagmatic relation of words" (p. 11). The study of collocations dates back to the 1960s as Hori (2004) illustrated. The concept of collocation is discussed in a seminar at Edinburgh University and among the participants were Halliday and Sinclair, whose paper "beginning of the study of Lexis" in 1966 shows some problematic issues related to the computational study of collocations.

Moreover, basic terms such as (node, span, cluster, collocation) have appeared since that time. The study of usual collocations in English is the major interest of both Halliday and Sinclair, whereas Angus McIntosh has an interest in studying unusual collocations in literary works. McIntosh's (1966) work "patterns and ranges" distinguishes between two notions range and pattern (Hori, 2004).

In the 1970s, one of the most influential works was *English Collocational Studies*, which applies the computer-based study of collocations, the authors of the work face problems with the size of the corpora and they have to cope with a certain problem related to the application of the computer. In the 1980s and onwards, there has been a dramatic development in technology mainly in computer and computational studies and the toolkits that are used in the analysis of texts as well as the compiling of data. All these have paved the way for fruitful studies of collocations. Important works include Sinclair's book (1991) *Corpus, Concordance, Collocation*. It is crucial to mention that all the scholars Halliday, Sinclair, McIntosh are originally influenced by Firth's (1957) ideas in his paper (Modes of Meaning). Based on what have been mentioned above, this paper aims at contributing to the field by examining collocation networks in practice.

### ***Studying Collocations in Literary Work: Investigating Authors' Style***

In this respect, Greenbaum (1970) emphasized that the study of collocation in literary language may reveal certain features of the writer's literary style. The idea of collocation networks was primarily suggested by Phillips (1983) in his Ph.D. thesis as Brezina, McEnery, and Wattam (2015) remarkably noted. This was followed by other important studies by Philips (1985, 1989) focusing on the same domain.

The years following show the interest of scholars to delve into this particular area of study. For example, Brezina et al. (2015) offered a new perspective on the study of the collocational networks by introducing GraphColl "graphical collocations" tool to examine and get insightful results about the semantic relations of words.

The work conducted by Brezina et al. (2015) is replicable of McEnery's study (2006). Instead of applying WordSmith tools, they used GraphColl with the main aim of focusing on methodological practices in finding collocational associations. Furthermore, they present some limitations of previous studies on collocation networks. Starting with Phillip's research (1989) whose methodology is unreplicable and the tool that is used for the analysis is not available anymore. Another limitation is represented by William's work (1998) though, is it replicable but the research gives few details about the span, and most importantly, it does not consider directionality as a criterion for its methodological purposes. In this study, they argued the word *affair* can be found in Bank National Corpora (BNC) collocates with the word *love* and the latter collocates with words like (madly, unrequited, undying) in the following: Madly Love, unrequited love, and undying love.

The above three words are not connected with "affair". However, in their argument, Brezina et al. (2015) notified that the relationship between the word *affair* and the words that are collocated with *love* should not be neglected and researchers should pay heed to the core word and its connections in context. It is worth noting that corpus tools help to identify essential or

eligible collocations for studying. Williams (2001) points out the major difference between collocations and idioms. He illustrates that collocations are easy to understand and less vague.

### ***Types of Collocations***

As a matter of fact, there are two types of collocations: lexical and grammatical collocations. The former type usually contains two lexical elements ( noun+ adjective, verb+ adjective). While the latter, basically is formed by combining a verb and preposition (e.g. depend on) or adjective with a preposition such as (good at, ready for, bored with). (for more details see El-Dakhs, 2015; Lewis, 2000 ). Thus, the current paper looks forward to revealing such combinations of the selected three words.

### **Methods**

The corpora of this research consist of a collection of abstracts that have been chosen from two different academic journals. The first corpus (29377 tokens) is a collection of abstracts collected from The *Journal of College of Education for Women* issued quarterly from the College of Education for Women, University of Baghdad / Iraq. The abstracts are related to non-native speakers of the English language and were published in the period 2018 to 2020. The researchers must translate the abstracts into the English language. The specialty of these research papers is in the field of Humanities including but not limited to (Arabic, English, History, Geography, Home Economics, Kindergarten studies, Educational and Psychological Sciences, Quranic sciences, sociology Computer sciences). The purpose behind examining these translated abstracts is to find the main collocational networks of mentioned nodes and their contexts. It is hypothesized that these words have been used similarly in comparison to native speakers.

While the second corpus consists of abstracts collected from another open access international Scopus indexed journal *Cogent Arts & Humanities* deals with arts and humanities and multidisciplinary areas of studies and accepts publications from different researchers all over the world. This corpus focuses on abstracts of research articles published during 2018 and 2020. The total size of this corpus is 25464 tokens and 5131-word types. To ensure objective results, the sizes, as well as the years of publications of the two corpora, are somehow close. However, the authors of the second journal are from different areas and linguistic backgrounds. This matter helps to check how a variety of authors around the globe employ word collocations in their academic writing.

### **Procedure and Data Analysis**

#### ***Extracting collocation networks Using GraphColl***

GraphColl is one of the tools in the free software #LancBox introduced by Brezina, *et al* (2015). It is a friendly tool that helps to identify collocation, keywords, concordances, frequencies, and dispersions. This can be systematically conducted by employing quantitative and qualitative methodologies. Updates to the functionality of this tool are continuous which makes this tool practical and friendly to use.

**Discussion and Results**

It has been noticed that the authors of these abstracts have used three words ( study, paper, research) interchangeably. Generally, these nodes are used to refer to academic writing work that presents an analysis, evaluation, and results. Their academic achievement is supposed to be scientific, rigorous, and objective. It is hypothesized in this current study, that the three words have shared collocations since they are used and employed constantly in academic contexts. Collocational associations can be pointed out systematically through GraphColl. This is what distinguishes this tool from the other myriad tools (e.g. AntConc, WordSmith Tools) with the same purpose of identifying collocations.

Shared collocations among the target nodes *study*, *paper*, and *research* can help researchers and those who are inquisitive in meaning, semantic relations, and the co-occurrence of words. From this theoretical standpoint, the following research questions have been tackled:

- 1- How do the three words study, paper and research are used interchangeably or synonymously?
- 2- What are the shared collocational associations surrounding these words?

GraphColl as a tool is used to identify the collocations. The statistical measure is the MI (Mutual information) score with statistic cut off value three) to discard infrequent associations. Other collocation parameters notation (CPN) include minimum collocate frequency C.5 and minimum collocation frequency equals NC.1. Function words removed.

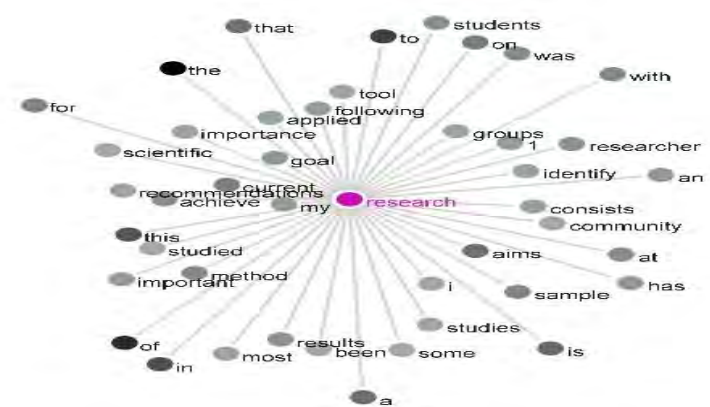


Figure 1. Collocation network around *research* [MI(3), C5, NC1, 5L 5R1]



Figure 2. Collocation networks around *research* [MI(3), C5, NC1, 5L 5R1]

In this figure, the statistical measure is the MI score with statistic cut-off value increased to five the results show that the most frequent words according to the frequency include ( this, aims, current, method, achieve, goal). These collocations show how the writers of the examined research papers aim to achieve certain goals by using verified methods. The distal demonstrative function word ‘this’ occurs 39 times in a collocational relationship with research and its frequency in the whole corpus is 195 which equals 20%.

Table 1. *Research collocations threshold 5 collocation frequency 5*  
 Freq:176 –collocates:19

index	position	Collocate	Freq coll.	Freq(corpus)
1.	L	My	7	7
2.	L	Goal	8	12
3.	L	Current	14	25
4.	R	Aims	21	52
5.	R	Groups	6	16
6.	M	Applied	6	17
7.	R	Following	7	20
8.	R	I	5	15
9.	R	Consists	6	19
10.	R	Identify	6	19
11.	L	Achieve	10	32
12.	R	Tool	5	19
13.	L	Method	11	44
14.	L	Importance	6	24
15.	R	Community	5	20
16.	L	Studied	5	22
17.	L	recommendations	6	29
18.	L	This	39	195

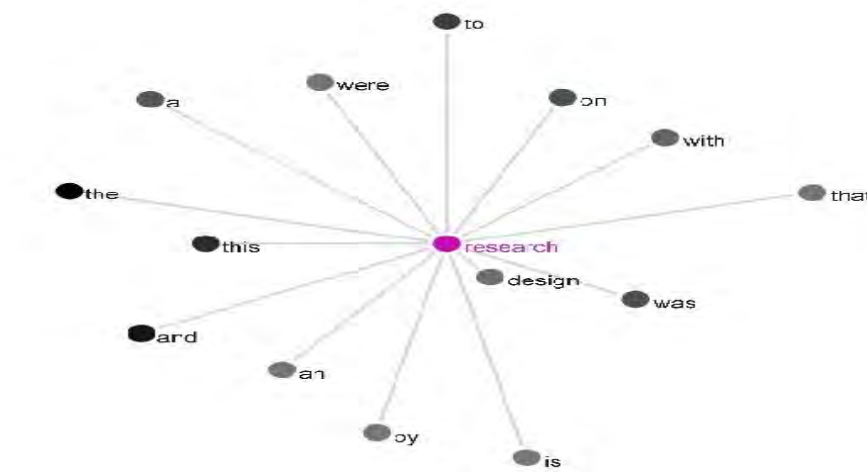


Figure 3. Collocation networks around *research* in corpus 2 [MI(3), C5, NC1, 5L 5R1]

Table 2. Research collocations threshold 5 collocation frequency 5

research						
Freq: 48 - Collocates: 14						
Index	Status	Position	Collocate	Stat	Freq (coll.)	Freq (corpus)
1	=	R	design	7.24113191...	6	21
2	=	R	was	5.45155173...	10	121
3	=	L	this	5.04848686...	17	272
4	=	R	an	4.75570514...	5	98
5	=	R	on	4.72655877...	10	200
6	=	R	with	4.56122104...	7	157
7	=	R	were	4.38173027...	5	127
8	=	R	by	4.07579422...	5	157
9	=	L	and	3.78952832...	24	919
10	=	R	is	3.53752495...	5	228
11	=	R	to	3.47513099...	13	619
12	=	R	that	3.32602084...	5	264
13	=	R	a	3.27883266...	9	491
14	=	L	the	3.17739699...	34	1990

From figure 3. and Table 2., it is clear that the word *design* as a content word has a strong collocational relation with the target node *research*. More contextual examples of the case include:

Occurrences	6/48 (2.36)	Texts	Corpus	Cogent Arts abstracts	Context	7	Display Text
Left	Node	Right					
a slow and processional approach to co-developing	research	and design questions, tie our process to					
framework connecting transformation design and participatory action	research	and point toward other potentially valuable frameworks					
complexity. To this end, using an experimental	research	design, two language institutes were randomly selected					
complexity. To this end, using an experimental	research	design, two language institutes were randomly selected					
m-learning for their accounting education. The survey	research	design was employed with copies of a					
concerning these lecture comprehension challenges. A qualitative	research	design was employed to collect data through					

Figure 4. the node 'design' mostly occurs in the right context

The node 'design' mostly occurs in the right context following *research* as shown in Figure 4. It is worth noting that function words have the biggest opportunity to accompany *research* in the second corpus, for instance, (was, this, an, on, with, were, by, and, is, to, that, a, the).

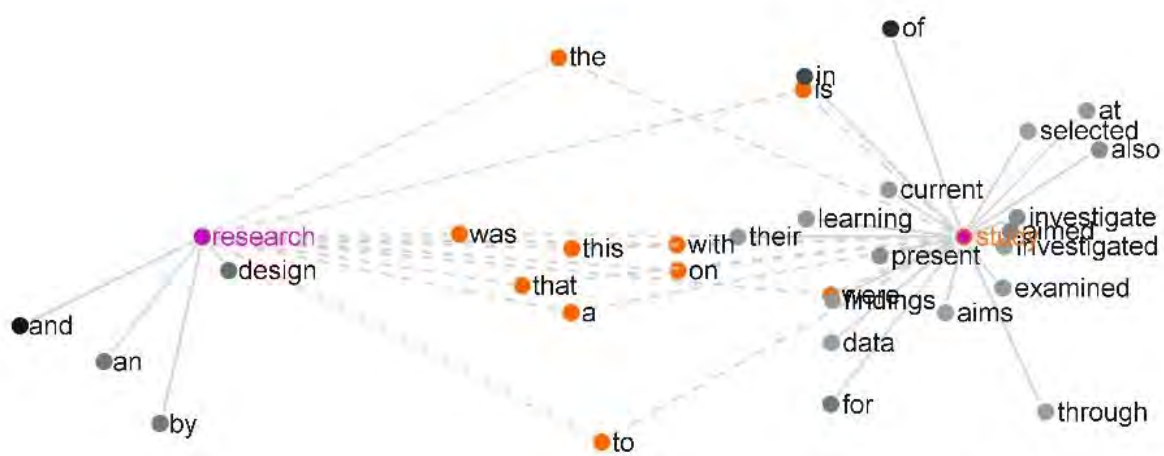


Figure 5. Shared collocations of the nodes research/study

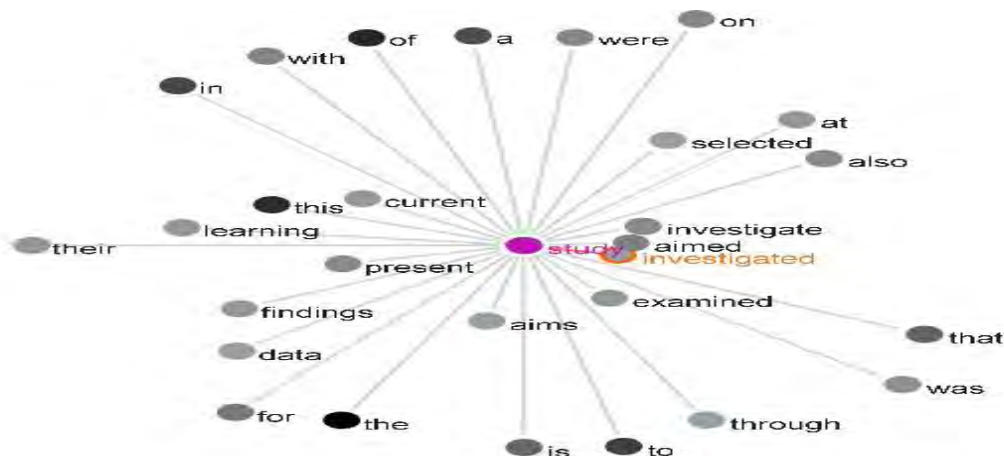


Figure 6. ‘Study’ collocates in Cogent Arts abstract corpus 2 threshold value 3 MI3

In a pursuit to find shared collocations of the nodes *study*, *research*, and *paper* in the two selected corpora ( abstracts of two journals). It is important to check the collocation of each target word. Figure 6. illustrates how the word *study* is surrounded with important content words opposite to *research* which inclines to co-occur with function words.

For more details about ‘*study*’ which has the total frequency in the corpus 140 and (28) collocates as shown in figure 6. and its surrounding relations with words focusing mainly on content words.

Table 3. Rank collocates Freq coll. Freq corpus position

1	investigated	5	6	R
2	aimed	10	13	R
3	investigate	8	12	R
4	examined	6	10	R
5	aims	5	10	R
6	present	8	18	L
7	current	6	14	L
8	this	69	272	L
9	selected	5	20	R
10	findings	6	33	L
11	learning	6	40	L
12	also	8	57	R
13	data	5	39	L

Table 3. reveals that writers use past tense form in writing their abstracts. It also illustrates that the associated words with the node *study*, such as (investigated, aimed, examined) in comparison with the present tense form (i.e. investigate, aims) as verbs in the right position as shown below in the examples in Table 4:









In this figure 10, the focus is mainly on basic words that function as shared collocation between *research* and *study*. The node *paper* seems to be isolated and shares only one word in for example ‘this research paper aims’. This leads to the fact that both nodes ‘study’ and ‘research’ are used interchangeably by the writers of published journal articles. They do not usually tend to use the word ‘paper’ in their abstracts.

Table 7. The twenty-one shared collocated among paper, study and research

ID	▲ Collocate	Freq (corpus)	No of nodes	Nodes
1	a	424	2	research, study
2	aims	52	2	research, study
3	an	91	2	research, study
4	at	90	2	research, study
5	current	25	2	research, study
6	has	98	2	research, study
7	identity	19	2	research, study
8	in	848	2	paper, research
9	is	354	2	research, study
10	of	1702	2	research, study
11	on	182	2	research, study
12	recommendations	29	2	research, study
13	results	68	2	research, study
14	sample	67	2	research, study
15	students	151	2	research, study
16	that	305	2	research, study
17	the	3054	3	paper, research, study
18	this	195	3	paper, research, study
19	to	730	3	paper, research, study
20	was	134	2	research, study
21	with	199	2	research, study

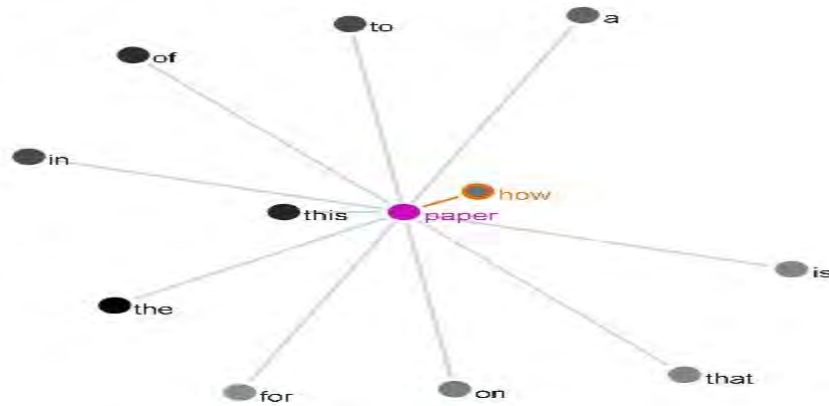


Figure 11 Study collocates in Cogent Arts abstract corpus

Index	File	▲ Left	Node	Right
3	Cognet arts a	external spaces in postcolonial contexts. The present	paper	attempts to explore the application of what
7	Cognet arts a	health crises, via TV documentaries. The present	paper	attempts to investigate J. M. Coetzee's Waiting
53	Cognet arts a	important part of the world. Furthermore, the	paper	aims to raise awareness of the fact
55	Cognet arts a	Prehistoric Scandinavian symbolic structures overall. The	paper	seeks to establish the role of religion
6	Cognet arts a	is the Dodol Village (Kampong Dodol). This	paper	intends to address the declining identity of
19	Cognet arts a	adaptation theory as its theoretical framework, this	paper	attempts to delineate how, in terms of
16	Cognet arts a	Hegelism, Platonism, and De Stijl. Hence, this	paper	aims to use Mondrian's writings, as a
15	Cognet arts a	in discourse theories, the authors of this	paper	analyze Philida to provide a new reading
48	Cognet arts a	of each form the objective of this	paper	is to provide a deeper and more
10	Cognet arts a	a long way. The objective of this	paper	is to analyse the factors that influence
8	Cognet arts a	the two cultures The task of this	paper	is to draw the possible new trajectories
50	Cognet arts a	to cognitive linguistics and literary studies. This	paper	aims to enter into the contemporary debate

Figure 12. Collocates of paper in Corpus 2

It is clear that the word ‘paper’ is used here by the authors to refer to study or research. Figure 12 shows there is a tendency that the node *paper* collocated with (to-infinitive) nominal clauses (the base form of the verb preceded by ‘to’) for instance:

- The paper attempts to investigate
- The paper aims to raise
- To establish
- To address

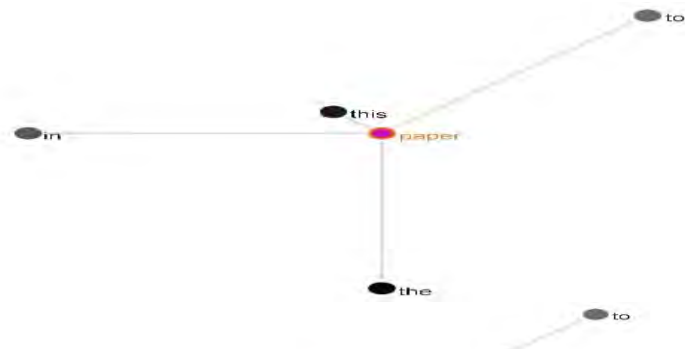


Figure 13. Collocates of the node *paper* in corpus 1

Search paper		Occurrences 22 (7.49)	Texts 1	▼ Corpus	Corpus 1	▼ Context 7	▼ Display Text
Index	File	Left	Node	Right			
1	Abstracts coll	silences than in the actual dialogue. This	paper	deals with Pinter's The Caretaker and how			
2	Abstracts coll	became weak in that era. The present	paper	tackles the Assyrian-Egyptian conflict in the 2nd			
3	Abstracts coll	this family with the Assyrian rulers. Also, this	paper	reveals the historical hostility between the two			
4	Abstracts coll	aid B.A. students in making their research	paper	more coherent, teaching this topic within Essay			
5	Abstracts coll	the main findings of the study. This	paper	aims to study Antonia's character and her			
6	Abstracts coll	pupils is not balanced. The present research	paper	aims at describing this phenomenon through statistics			
7	Abstracts coll	semantics of the English preposition in. The	paper	pays attention to the polysemous words Harry			
8	Abstracts coll	selected to be the data of this	paper	The first is an Arabic one by			
9	Abstracts coll	of the article it refers to. This	paper	aims at examining this hypothesis in relation			
10	Abstracts coll	by a number of news agencies. The	paper	follows Halliday (1967) information structure theory by			
11	Abstracts coll	is the most appropriate one. In this	paper	I examined the specification phenomenon (المفهوم) in			
12	Abstracts coll	the part of the qualitative analysis, the	paper	will explain some related concepts generated in			
13	Abstracts coll	various material and non-material means. This	paper	sheds light on the concept of the			
14	Abstracts coll	for themselves nor for their nation. This	paper	is important as it discusses the post-colonial			
15	Abstracts coll	spiritual loss of these young natives. The	paper	aims at examining this state during that			
16	Abstracts coll	that period, and the findings of the	paper	show that the Cherokee nation has no			
17	Abstracts coll	W3, Salinity, C3 and C2. This research	paper	attempts to explore problems facing the teaching			
18	Abstracts coll	critical kind of research. As such, this	paper	is a critical pragmatic investigation for this			
19	Abstracts coll	underpin the illegitimate issue under investigation. This	paper	sets for itself the task of answering			
20	Abstracts coll	phenomena common among kindergarten children. The current	paper	highlighted an important character represented by the			
21	Abstracts coll	tenth American president. The importance of this	paper	is by revealing his personality in terms			
22	Abstracts coll	other methods of all this? In this	paper	we try to show the importance of			

Figure 14. Concordance lines of the node *paper* in corpus 1

Interestingly, the concordance lines of the word *paper* shows that the writers in corpus 1 use the language differently in terms of using verbs directly after the node *paper* as shown in the right context instead of to-infinitive for instance:

- This paper reveals
- The present paper tackles
- This paper deals with
- This paper (sheds light on, attempts, aims, shows, highlighted, follows...etc).

## Findings

Though as researchers we might think that the nodes (research, paper, and study) are used interchangeably in the corpora, quantitative corpus data shows these words are not interchangeable at least within the limits of the current study. The data point to the fact that both *paper* and *study* share different collocates. In contrast, findings indicate that both nodes ‘study’ and ‘research’ are used interchangeably by the writers of published journal articles. They do not tend to use the word ‘paper’ in their abstracts. This implies that ‘study’ and ‘research’ have shared collocational networks as illustrated in *figure 9*.

The findings of this current study have important significance to researchers/ teachers who are interested in semantic relations and the study of collocational networks. Furthermore, it can give foreign learners of language an idea about the associations of words in context. This emphasis is pointed out in the work of Brezina et al. (2015) “.. collocational relationship has important implications for our understanding of language and word meaning. Collocation networks show how meanings of words are formed through multiple repeated associations that can be documented only in language corpora” (p. 165). Moreover, this study shows teachers a way of engaging students or learners to the target language they learn by employing a corpus tool to facilitate locating aspects that can be missed by manual analysis only.

## Conclusions and Recommendations

The research is primarily conducted with aim of examining the collocation network of selected words in the abstracts of published papers in two respected international academic journals. It is worth mentioning here that corpus linguistics helps to find ways to examine and study collocations in context. GraphColl as free software helps researchers to gain a better understanding of some unexplored aspects of authentic language. Introducing a new dimension as Brezina et al. (2015) emphasized to the other dimensions adds reliability and systematicity to the investigation of collocations. The findings show that ‘research’ and ‘study’ are used rather interchangeably in the writing of the abstracts of the two corpora; however, ‘paper’ has fewer shared collocations in the same academic context. Moreover, the nodes have mainly lexical collocations instead of grammatical ones.

Finally, within the teaching context, the importance of using corpora in the classroom lies in the main concept of introducing students to discover patterns of the language and to enhance students' linguistic competence through noticing and dealing with authentic language. They can learn about the behavior of words in context. The relationship between words, thus, is revealed through the collocational networks. However, some limitations might have appeared. Some teachers might feel intimidated by using technology to teach language. They might lack the required experience to deal with computational tools. Moreover, the lack of the necessary equipment is another possibility that teachers might face. One more issue, a large number of students can hinder learning and require systematic guidance and instructions. However, these issues can be tackled by the use of free online corpora and other friendly tools that can be accessed by the mobile devices of the learners. Finally, the integration of online teaching through many available platforms can save time and give more practice opportunities.

## About the Author:

Eman Adil Jaafar is Assistant Professor of linguistics at the Department of English/ College of Education for Women / University of Baghdad. Her research interests include but are not restricted to stylistics, corpus and cognitive stylistics, and applied linguistics. <https://orcid.org/0000-0002-8880-8930>

**References**

- Baker, P. (2016). The shapes of collocation. *International Journal of Corpus Linguistics*, 21(2), 139-164. <https://doi.org/10.1075/ijcl.21.2.01bak>
- Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: A new perspective on collocation networks. *International Journal of Corpus Linguistics*, 20(2), 139-173.
- Brezina, V. (2016). Collocation networks: exploring associations in discourse. In P. Baker, & J. Egbert, (eds.), *Triangulating Methodological Approaches in Corpus Linguistic Research* (pp. 102-119). Routledge: London.
- Brezina, V. (2018). *Statistics in Corpus Linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- Davis, P., & Kryszewska, H. (2012). *The company words keep: Lexical chunks in language teaching*. Delta Publication. <https://p302.zlibcdn.com/dtoken/5f37b00c856083386ec250011ad80ef4>
- El-Dakhs, D. A. (2015). Collocational Competence in English Language Teaching: An Overview. *Arab World English Journal*, 8 (1). DOI: <https://dx.doi.org/10.24093/awej/vol6no1.5>
- Firth, J. R. (1957). 'Modes of Meaning', in J. R. Firth, *Papers in Linguistics, 1934–51*. London, New York: Oxford University Press.
- Greenbaum, S. (1970). *Verb-intensifier collocations in English: An experimental approach* (Vol. 86). Walter de Gruyter.
- Gablasova, D., Brezina, V., & McEnery, T. (2017). Exploring learner language through corpora: comparing and interpreting corpus frequency information. *Language Learning*, 67 (S1), 130–154. <https://doi.org/10.1111/lang.12226>
- Halliday, M. A. K. (1966). Lexis as a linguistic level. In C. E. Bazell, J. C. Catford, M. A. K. Halliday, R. H. Robins. (Eds.), *In Memory of J. R. Firth* (pp. 148–162). London, Longman.
- Hori, M. (2004). *Investigating Dickens' style: A collocational analysis*. Basingstoke and New York: Palgrave Macmillan.
- Lewis, M. (2000). Language in the lexical approach. In M. Lewis (Ed.). *Teaching collocation: Further developments in the lexical approach* (pp. 133-134). London: Language Teaching Publications.
- McIntosh, A. (1966). Patterns and ranges. In A. McIntosh, & M.A.K. Halliday, (eds.) *Patterns of Language: Papers in General Descriptive and Applied Linguistics* (pp. 183-199). London: Longman.
- McEnery, T. (2006). *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Abington, UK: Routledge.
- Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14). Amsterdam: John Benjamins Publishing.
- Nagano, T., & Kitao, K. (2007). Using Simple Computational Linguistic Techniques for Teaching Collocations. *Journal of Culture and Information Science*, 2(1), 1–15.
- Phillips, M. K. (1983). *Lexical macrostructure in science text*, (Unpublished doctoral dissertation). University of Birmingham, Birmingham, UK.
- Phillips, M. (1985). *Aspects of Text Structure: An Investigation of the Lexical Organisation of Text*. Amsterdam, Netherlands: North-Holland.

- Phillips, M. (1989). *Lexical Structure of Text* [Discourse Analysis Monograph 12]. Birmingham, UK : University of Birmingham
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2004) *Reading Concordances: An Introduction*. London: Pearson.
- Williams, G. (1998). Collocational networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics*, 3(1), 151–171. DOI: 10.1075/ijcl.3.1.07wil
- Williams, G. (2001). Mediating between lexis and texts: Collocational networks in specialised corpora. *ASp. la revue du GERAS*, (31-33), 63-76. <http://dx.doi.org/10.4000/asp.1782>