Research Paper

Journal of Language and Translation
Volume 14, Number 4, 2024, (pp.121-136)

Collocations and Colligations in the Scientific Research Articles in Terms of Lexical Priming Theory: Analysis of Hard and Soft Sciences

Leila Ghemi[1], Nesa Nabifar[2*], Davud Kuhi[3]

[1]Ph.D. Candidate, Department of English, Tabriz Branch, Islamic Azad University, Tabriz, Iran
[2*]Assistant Professor, Department of English, Tabriz Branch, Islamic Azad University, Tabriz, Iran
[3]Assistant Professor, English Language Department, Maragheh Branch, Islamic Azad University, Maragheh, Iran

Received: June 24, 2023 Accepted: September 18, 2023

Abstract

Lexical Priming Theory claims that the collocational priming of a word depend on the genre or discipline in which it occurs. In this study, attempts were made to compare the collocates and colligates and the positions of the collocates prior and subsequent to the node word in two specialized corpora of soft science and hard science to see if they are realized in the same or different ways. To serve this purpose, the introduction section of 1000 RAs of two disciplines of soft and hard were gathered from disciplines of applied linguistics, sociology, and psychology (soft science) and computer science, chemistry, physics and medical science (hard science) in order to find a relationship between these two types of RAs with differences in discipline and the authors' use of collocations and colligations. The findings revealed that the frequency of collocations that employed in the introduction sections of soft and hard sciences were not the same and the researchers of soft science in designing their research articles' introduction sections employed more collocations in comparison to the researchers in hard science. Hence, based on the results of Chi-square data analysis, it can be claimed that there is a significant difference between two groups of authors in using colocations. However, in terms of colligations, both groups of authors used colligations with the same rates and similarities in the right and the left positions of the collocations, hence it can be claimed that there is no difference in two sciences in terms of high dominant colligations.

Keywords: Colligations; Collocations; Introduction Sections; Hard and Soft Disciplines

INTRODUCTION

A novel theory of language as well as the lexicon is the Lexical Priming Theory (Hoey, 2005). According to this view, lexis and grammar play different roles, with lexis having a sophisticated and systematic structure and grammar emerging from this lexical structure (Patterson, 2016). In an effort to explain how naturalness is attained and how an explanation of what is natural may affect explanations of what is possible, this new theory attempts to provide examples. Collocation is a crucial aspect of naturalness in this context (Hoey, 2012). Lexical priming is, as its title boldly claims, a new theory; not because it prioritizes lexis, but because it prioritizes the individual's experience of language acquisition and use, taking lexis as a starting point. Priming is viewed here as a bottom-up process of lexical pattern- forming, expressed in terms of phraseology, not grammar (Philip, 2009). The rejection of grammar as an Priming, based on this view, are psychological phenomena and each person's experience of any particular word is inevitably unique (Novakova & Siepmann, 2020). It follows then that the priming for a word may vary from person to person, being based on different

*****Corresponding Author's Email: *n.nabifar@iaut.ac.ir*

encounters with the word in different contexts. According to this view, as we encounter and acquire a word or phrase through repeated encounters (spoken or written), we at the same time subconsciously note the contexts in which we are encountering the expression in question; we are primed by the encounter (Almela-Sánchez & Cantos-Gómez, 2019). In a way, we build up a concordance of the word or group in our heads and process it. When called upon to use the word or phrase, we draw without thinking upon this concordance knowledge and in so doing we simultaneously reinforce our own priming and those of our listeners or readers (Hoey, 2012).

In addition to the collocation, Hoey's theory of lexical priming encompasses the other term called colligation. The term colligation is used to refer to recurrent combinations of lexis and grammar (Thompson, 2014; Tongnini-Bonelli 2001). These concepts highlight the fact that the choice of vocabulary is not free but regulated by constraints on word co-occurrence. Moreover, evidences have shown that types of words grammatical categories favored or avoided by particular word sense vary considerably according to contextual usage and language variety. Despite the independent development of the term in several places, it has developed the same sense, namely the grammatical associations that a word forms with its environment or the grammatical pattern in which it participates.

To date, as stated, the most extensive treatment of the concept colligation has been presented by Hoey (2005). In Hoey's theory of lexical priming (i.e., a statistically based theory of linguistic competence), colligations play a crucial part in what it means to know a language. According to Hoey (2005, p. 43) colligation actually encompasses three distinct aspects of distributional attraction between linguistic items: (i) the relationship between a lexical item and a grammatical context (e.g. [consequence + B E + subordinate clause]; Hoey, 2005, pp. 57–58), (ii) the relationship between a lexical item and a particular syntactic function in which the item can be used (e.g. consequence is often used as part of a complement; Hoey, 2005, pp. 44–48), and (iii) the relationship between a lexical item and the position in a phrase, clause, sentence, text or discourse where the item can be used (e.g. consequence is often used as part of the theme in a sentence; Hoey 2005, pp. 49–52). Thus, Hoey uses colligation as a cover term which encompasses both grammatical patterns and patterns of information structure associated with a lexical item. It is important to note that all of the relationships above can be positive as well as negative, i.e., lexical items are primed to co-occur with some grammatical features while they are also primed to avoid others.

Following Pecorari (2006), English for Specific Purposes (ESP) research findings in different disciplines and genres must cover not only what features occur, but also how and why they are used. The first of these questions, what linguistic features characterize a given genre within a given subject area, is most directly answered by corpus approaches. The importance of genre knowledge in helping language learners to understand and master academic, professional or educational discourse has widely been acknowledged for over two decades (Swales, 2004). Generic knowledge is one of the competencies needed (Allen, 1989). As most of our everyday knowledge, genre knowledge is generally tacit and would be hard for most readers to articulate as any type of comprehensive and coherent framework. Obviously, one requires to encounter adequate instances of a genre so as to distinguish shared features as being characteristic of it (Swales, 2004). The uniqueness of features of genres can be extended to individual languages as well. Kaplan (1966) posited that each language has its rhetorical patterns, which can bring about variation in rhetorical patterning of a specific type of genre. Unfortunately, the genre approach in corpus work has merely employed a macro-analysis perspective focusing on the rhetorical structures of texts and has overlooked the micro-analysis of lexical patterns of texts via lexical priming models which form a significant and indispensable part of any genre due to the fact that the same lexical items might be patterned and used differently across different genres and disciplines. To the best of our knowledge, however, there have been very few analyses of the distribution and patterning

behavior of collocations and colligations across disciplines. Therefore, the present study aimed to scrutinize on the terms such as lexical priming, collocations and colligations in order to investigate lexical priming in introduction section of two types of genres that is hard versus soft sciences and look for the patterns existed in using collocations and colligations by the authors of these disciplines.

LITERATURE REVIEW

Lexical Priming

Hoey's (2005) priming draws on psycholinguistic arguments and it claims that as the word is learnt through encounters with it in speech and writing, it is loaded with the cumulative effects of those encounters such that it is part of our knowledge of the word that it co-occurs with other words. Hoey (2005) considers priming as the most appropriate psychological concept to account for collocation. He believes that if words were stored in our minds separately or in sets, collocational naturalness would be inexplicable (Pace-Sigge, 2018). Hoey (2012) resorts to the notion of semantic priming to discuss the way a priming word may provoke a particular target word. For example, previously given the word body, a listener will recognize the word heart more quickly than if they had previously been given an unrelated word such as trick; in this sense, body primes the listener for heart. The word body sets up a word association with heart, which the word trick does not (Alipour & Biria, 2015). Hoey (2005) contends that the theory grew out of an increasing awareness that traditional views of the vocabulary of English were out of kilter with the facts about lexical items that are routinely being thrown up by corpus investigations of text. As Dong and Buckingham (2018) believed, a key aspect of recurring linguistic features is the association between words or phrases and their textual positions. Lexical items, according to Hoey (2005, p. 13) "are primed to occur in or avoid, certain positions within the discourse". An analysis of textual colligation, the term Hoey (2005) uses to denote such priming, explores the textual position of linguistic markers in relation to textual structures, and may also examine the interaction between the textual position and discourse functions (Barlow, 2016).

Collocation

As stated, priming was considered as the most appropriate psychological concept to account for collocation (Hoey, 2012). He believes that if words were stored in our minds separately or in sets, collocational naturalness would be baffling (Römer, 2022). In other words, as Bartsch (2004, p.5) believed by pointing out that "you shall know a word by the company it keeps". Furthermore, with the theory of priming, Hoey (2005) offers an explanation for the existence of collocation: "the only explanation that seems to account for the existence of collocation is that each lexical item is primed for collocational use" (p. 386). Collocation has been studied for at least a half-century. Firth (1957) was the first who introduced the notion of collocation into contemporary linguistics and it is believed that Firth's collocation idea is essentially quantitative (Krishnamurthy, 2000). Although many scholars have introduced different definitions of collocations, never have they arrived at a consensus (Mel'cuk, 1998; Wray, 2002; Nesselhauf, 2005; Xu, 2018). However, one important definition has been that of Benson et al. (2010), who referred to collocations as "fixed, identifiable, non-idiomatic" combinations which are used repetitively in a language (p. 19). Based on the definitions, linguists have recognized for some time that words in natural language are neither randomly combined into phrases and sentences nor that they are only constrained by the rules of syntax (Hunston, 2007). Curiously, this basic fact about collocations and, at the same time, their rather diverse and apparently idiosyncratic behavior, has been taken out of focus by a substantial part of contemporary mainstream linguistics which has been primarily concerned with examining language from a theoretical perspective (Subtirelu & Baker, 2017). In particular, generative linguistics in the Chomskyan tradition demotes all lexical and syntactic idiosyncrasies safely into the realm of the lexicon (McEnery & Wilson, 2001).

Colligation

Colligation is the concept proposed by Firth (1957) to refer to "the interrelation of grammatical categories in syntactical structure" (p. 12).

It also is used to refer to recurrent combinations of lexis and grammar (Tongnini-Bonelli 2001). These concepts highlight the fact that the choice of vocabulary is not free but regulated by constraints on word co-occurrence. Moreover, evidences have shown that types of words grammatical categories favored or avoided by particular word sense vary considerably according to contextual usage and language variety. Despite the independent development of the term in several places, it has developed the samesense, namely the grammatical associations that a word forms with its environmentor the grammatical pattern in which it participates (Dahunsi & Ewata, 2022). Halliday (1994) formulates the colligational relationship in terms of sentential position, this extension which means that colligation may be interpreted as going beyond traditional grammatical relations and embracing such phenomena as the positioning of a word or word sequence within the sentence or paragraph and even its positioning within the text as a whole; a colligational statement can be negative as well as positive. So, it is a logical colligational statement to say of a particular lexical verb that it does not occur with the main auxiliaries or that it avoids sentence-final position (Gu, 2017).

These years have experienced a great number of researches dedicated to lexical priming in general and collocations and colligations in particular, for instance, Concerning whether colligations of the same word differ in different disciplines, Haghverdi, Biriya, and Alipour (2013) compared the colligations of one noun (group) based on Hoey's (2005) framework across two specialized corpora of hard and soft sciences to find out whether they are realized similarly or differently in different disciplines. Results revealed that in different disciplines noun groups colligated with different grammatical structures and functions. In other words, nouns tend to occur in or avoid certain position in different disciplines.

Dahunsi and Ewata (2022) examined lexical bundles, a type of multi-word expressions, to understand their structure and co-occurrence possibilities with other syntactic elements. Two items of lexical analysis software were used to extract three-word lexical bundles with a minimum of 50 frequencies per corpus. The syntactic structures of the identified lexical bundles were determined, and their in-corpus usages were analyzed for their colligational characteristics. Results showed that both corpora had instances of general and genre-specific lexical bundles (LBs) with varying frequencies. Five categories of lexical bundles with different structural patterns and peculiar colligational characteristics were identified in the study.

Moreover, the studies of words and collocations have been widely considered and conducted in various disciplinaries (Stella, 2015; Thongvitit & Thumawongsa, 2017; Trinant & Yodkamlue, 2019; Yanxia & Weicai, 2017). For example, Thongvitit and Thumawongsa (2017) aimed to examine the types and count frequency of grammatical and lexical English collocations used in the abstracts of research articles in the field of liberal arts and humanities, which were written by Thai EFL writers from 2010 to 2015, and to identify misused English collocations produced by the Thai EFL writers. It was found that noun + preposition and adjective + noun collocations were grammatical and lexical collocations that were used the most often. The most often misused grammatical and lexical collocations found in the study were noun + preposition and verb + noun collocations. Also, in a comparative study, Yanxia and Weicai (2017) used corpora to compare linguistics and medical academic English lexical chunks and analyze their similarities and differences in lexical structure fixation, structural form, and discourse function to reveal the usage pattern of academic lexical chunks. The results showed that 1) linguistics and medical scholars prefer to use more productively semi-fixed lexical chunks, 2) there are significant differences in the species and frequency of the two most used lexical chunks in the two corpora, and 3) thehigh frequency four-word lexical chunks are similar in the structural form, but there are differences in the structure fixation. More recently, Suraprajit (2021) explored the frequent use of nouns, verbs, and adjectives together with the grammatical and lexical English collocations in the logistics magazines published in Thailand during 2019 – 2020. The findings listed that the most frequent function and content words such

as cargo and said as well as the high type of grammatical collocation such as noun + preposition and noun + verb.

It is worth noting that although there are a bulk of research in the filled of lexical priming including collocation, colligations, and word associations, however, there is a requirement for extra research as authors of research articles in different disciplines and even EFL learners still face some difficulties and problems in using the words and collocations (Boonraksa & Naisena, 2022; Dokchandra, 2019;). Therefore, by investigating the frequency of the collocations and the dominant patterns of collocations and colligations found in authentic usage such as scientific research articles, it could enhance authors and learners' awareness of these terms. That is because such information could present the real use of language that appears daily for specific purposes, especially when the meaning of each word could be varied when it appears in different disciplines. Therefore, studying the meaning of the word for the specific field should be given more attention in order that the learners could understand the real meaning in that specific context and gain more vocabulary knowledge which is varied in different disciplines. Due to the importance of collocations and colligations, the current study was an attempt to investigate the high frequent collocations and colligations as well as their dominant patterns in the introduction sections of scientific research articles authored by two sciences of hard and soft a corpus-based approach, which is still under-explored. The logic behind choosing Introduction sections of research articles is the fact that writing introductions, according to Swales and Feak (1994), is widely thought to be difficult and troublesome, and writing a solid introductory section typically appears to be a struggle for researchers, even though they are not novice. To this end, the following research questions were formulated:

RQ1: *How frequently are collocations and colligations employed in the introduction sections of soft and hard disciplines' research articles based on Hoey's (2005) lexical priming model?*

RQ2: *What are the differences in the use of collocations and colligations by the authors of soft and hard disciplines?*

RQ3: *What are the dominant patterns of collocations and colligations that the authors of soft and hard disciplines used in the introduction sections of research articles?*

RESEARCH METHOD

Corpus

The data for the present study was obtained from the introduction section of 1000 research articles (Ras) of two disciplines of soft and hard. The corpus of this study was gathered from disciplines of applied linguistics, and psychology (soft science) and computer science, physics and medical science (hard science) in order to find a relationship between these two types of RAs with differences in discipline and the authors' use of collocations and colligations. To ensure the generalizability of the results to the target discourse and account for representative practices of discourse community in different disciplines, leading journals in both hard and soft sciences were selected based on consultation with discipline experts and the Impact Factors (MIFs) ranging from 1 to 5 that were reported in the Journal Citation Reports (JCR) in 2015. Five data bases including Elsevier, Sage, Springer, Taylor & Francis, and Wiley Online Library comprise the sources of RAs in two disciplines published between 2010 and 2020. Research articles were randomly selected from each database, yielding a corpus of 1000 research articles of approximately 603,000 words (soft science comprising of 288,582 words) and (hard science comprising of 314,471 words). In the selection of this corpus, we ensured that there is a proportionate number of native and non-native writers.

Procedure

The design of this study is Corpus Corporation. In addition, the design was comparative in nature. The comparison was made between hard and soft sciences in the use of colligations and collocation based on Hoey's lexical priming framework (2005) and the results were analyzed quantitatively and descriptively. Also, the design was exploratory since as Dornyei (2005) states, the exploratory design is conducted about a research problem when there are few or no earlier studies to refer to. The focus is on

gaining insights and familiarity for later investigation or undertaken when problems are in a preliminary stage of an investigation. Exploratory research is flexible and can address research questions of all types (what, why, how).

To analyze the corpus, the introduction sections of two disciplines of soft and hard were identified in the articles and stored separately. It is important to note that these are slang phrases for comparing scientific subjects based on perceived methodological rigor, exactitude, and objectivity. Natural sciences are "hard," according to Wilson (2012), but social sciences are typically defined as "soft." The articles that comprised of isolated heading or sections such as introduction, literature review, method, and so were selected and those articles, which the introduction sections were embedded in the review section and vice versa were excluded from the analysis and converted into plain text files (txt). Afterwards, using AntConc software (Anthony, 2011), we identified the frequency of collocations in both disciplines. After that the necessary data gathered in collocations, it was analyzed by lexical priming model that accounts for collocation since as Hoey (2005) stated each term is mentally linked for collocational use. In accordance with the model, for analyzing the collocations, the noun group was selected to be compared for its collocational (the word or words that characteristically accompany a term) patterns across the corpora. After above phases, the rates and frequency of the collocations in both disciplines were presented in tables. The other focus of this study was on analyzing the data based on colligations. For exploring the colligations used in the corpora, the researcher scrutinized on the high frequently used colligations in both hard and soft sciences and after that she reported on the dominant patterns of colligations in the form of qualitative data analysis. To ensure the reliability of the analysis in the process of data categorization 10% of the data was rechecked and reanalyzed independently for collocations and colligations by a second researcher (a Ph.D. graduate of TEFL) who was briefed about the purpose of the study by the researcher. Also, the field of study of this expert was discourse analysis, and she was familiar with the data analysis phase. The second rater coded 10% of the data, taken randomly from the corpus and finally, the inter-rater reliability was estimated and reported. The inter-rater agreement, measured using Cohen's Kappa formula, was found to be Kappa = 0.929, $p = 0.000$.

Data Analysis

The current research as both qualitative and quantitative data analysis, tried to gain the aim of the study. The first research question was reported in terms of frequency and percentage to report on the rates of collocations and colligations, while to find the existence of any significant differences between two types of research articles in terms of the rates of collocations and colligations (the second research question), the results were compared via Chi-Square data analyses. In terms of the third research question, the dominant patterns of collocations and colligations in each corpus were investigated and it was compared with each other through descriptive data analysis.

RESULTS

In the present study, the corpora were analyzed in the same line with Hoey's (2005) model. In other words, we employed the same method to analyze the data. Therefore, the top high frequent words (both function words and content words) were detected in both disciplines. Table 1 shows the top high frequent words authors in soft and hard sciences used in their research articles.

Table 1

Top High Frequent Words in Soft and Hard Sciences

Rank	Soft/Types=15967/Tokens=286133		Hard/Types=17917/Tokens=311550	
	Frequency	Word	Frequency	Word
1	13197	the	17674	the
2	11088	of	12851	of
3	9151	and	10039	and
4	8021	in	8457	in
5	7272	to	7134	to
6	5237	a	2958	a
7	4769	‘s	4135	is
8	3254	that	3487	for
9	3066	is	2958	with
10	2731	as	2492	that
11	2469	for	2356	as
12	2236	on	2239	‘s
13	2164	language	2101	are
14	1893	this	1745	on
15	1717	are	1661	by
16	1700	with	1656	be
17	1690	l	1623	this
18	1583	writing	1549	have
19	1413	by	1457	or
20	1405	their	1283	been
21	1336	learning	1244	has
22	1310	research	844	patients
23	1284	have	777	studies
24	1214	students	776	study

Table 1 shows that the soft science authors with the types of 15967 and tokens of 286133 used both function and content words, very similar to those of hard science with the types of 17917 and tokens of 311550. Based on the results of data analysis, in both corpora, the function words such as “the, of, and, in, to, a, as, for” were used in high frequency. The five high frequent content words in soft science are *Language* with the frequency of *(F=2164)*, *Writing (F=*1583*)*, *Learning (F=*1336*)*, *Research (F=*1310*)*, and *Students (F=*1214*)* were the high frequent words after the function words. The number of words in soft science was 288,582 words with the total number of collocate types of 3317 and total number of collocate tokens of 21640. As it is clear crystal, from the first 24 high frequent words in the soft science, there were just five content words as mentioned above. To detailly analysis the rations of the collocates, the relative frequency ratio (1000) was reported. Table 2 shows the frequencies, percentages, and relative frequency ratio of collocations in soft science.

Table 2

Rates of Collocations (Content Words) in Soft Science

Collocations	Frequency	Percentage	Relative frequency ratio (1000)
Language	2164	0.75%	25.98
Writing	1583	0.55%	19.05
Learning	1336	0.46%	15.94
Research	1310	0.45%	15.59
Students	1214	0.42%	14.55

Based on the results obtained from Table 2, the relative frequency ratio (1000) of the collocate 'language', as the high frequent relative frequency ratio (1000), was 25.98 and the relative frequency ratio (1000) of 'student', as the low frequent relative frequency ratio (1000), was 14.55, and the relative frequency ratio (1000) of two collocates of 'learning' and 'research' were the same.

As already stated, the number of words in hard science was 314,471 and it is worth stating that the same procedure was conducted for the hard science corpus. However, the story was different in hard science because the number of the content words were less than the function words and the first 21 words 9See Table 1) were related to function ones. Moreover, the function words of 'was, it, et al, which, these, and can' were ignored in the list and the three content words of 'patients with the frequency of *(F=844)*, studies *(F=777)*, and study *(F=776)* were reported. Table 3 shows the frequencies, percentages, and relative frequency ratio of collocations in hard science

Table 3
Rates of Collocations (Content Words) in Hard Science

Collocations	Frequency	Percentage	Relative frequency ratio (1000)
Patients	844	0.27%	85.86
Studies	777	0.25%	79.50
Study	776	0.24%	76.32

Based on the results obtained from Table 3, the relative frequency ratio (1000) of the collocate 'patients', as the high frequent relative frequency ratio (1000), was 85.86 and the relative frequency ratio (1000) of 'study', as the low frequent relative frequency ratio (1000), was 76.32, and the relative frequency ratio (1000) of collocate of 'study' was between two collocates of 'patients' and 'studies. To investigate the existence of any significant difference between two corpora in terms of collocations, a Chi-square data analysis was run. Based on the results, the chi-square statistic is 44.1896. The *p*-value is < 0.00001. The result is significant at $p < .05$, then it can be claimed that there is a significant difference between two disciplines in terms of collocations based on Hoey's (2005) lexical priming model. In fact, the authors of research articles in soft science used more collocations in the introduction sections, in comparison to those in hard science.

As already stated, based on Hoey's model (2005), for analyzing the collocations, the noun group was selected to be compared for its collocational patterns across the corpora, to this end, function words were ignored in the current study and just the highly frequent nouns, no difference in their specificity or generality, were chosen from corpus. Based on the results of data analysis, in both corpora, the function words such as "the, of, and, in, to, a, as, for" were used in high frequency. In terms of the content words in soft science, the high frequent word of 'language' along the collocates were identified ana analyzed to explore the dominant pattern in colligations (Figure 1).

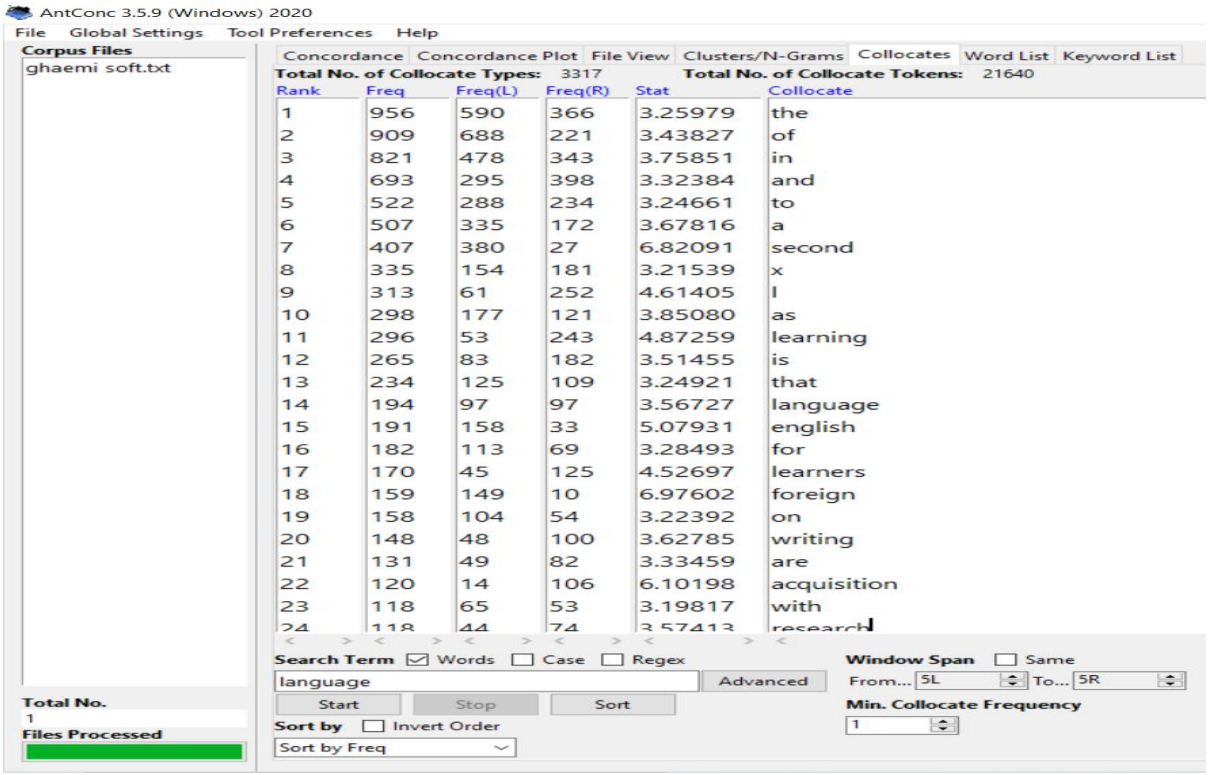

Figure 1

The top collocates of the noun 'language' occurring immediately prior and subsequent to the node word in the Soft Sciences

The results revealed that from among 15 top collocates of 'language', the content words of 'second, learning, language, English' were the high frequent ones that from among them, the high common positions of the words of 'second and English' were in left hand in comparison to the word of 'learning' that the least common position was in left hand rather that the right hand. Analyzing the collocates of 'language +second' in soft science corpus revealed extra valuable information (See Figure 2).

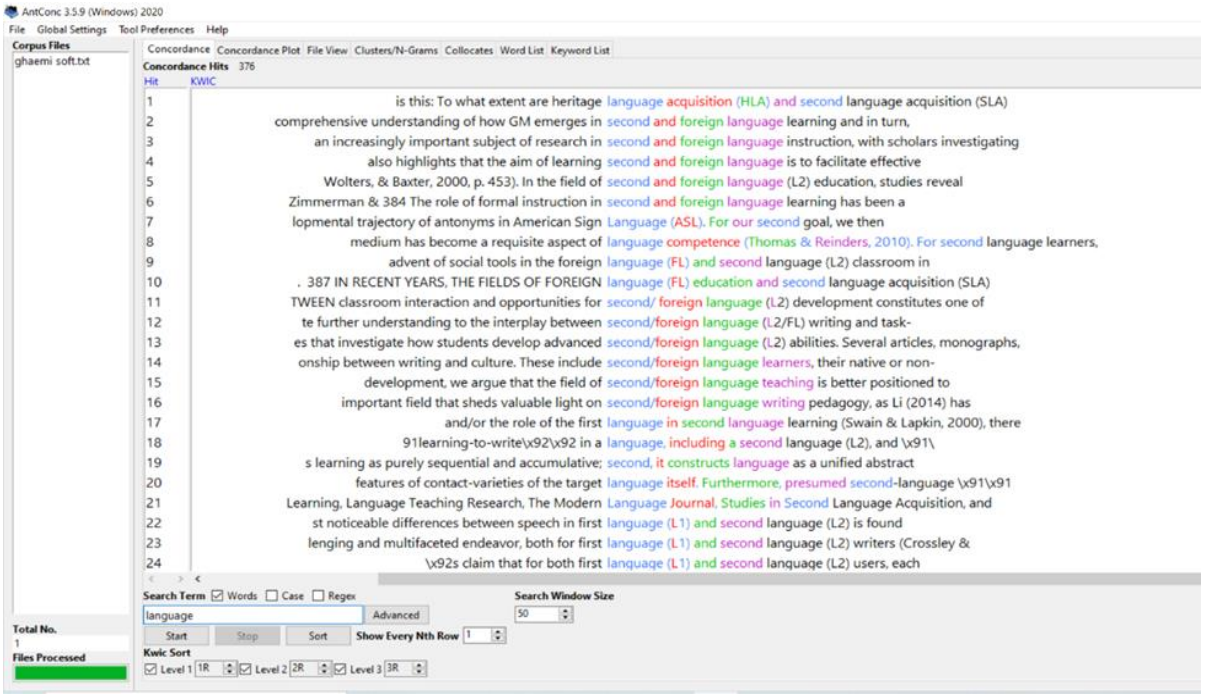

Figure 2

Instances of the collocate of 'language + second' in soft science

Based on the detailed analysis, the results showed that the collocate of 'language +second' in soft science are observed in both left hand and right hand, however, the weight of left hand (F=380) was higher than the right hand (F=27). It means that second can placed in the left hand of the content word of 'language' as the following three instances from the above list (Figure 2) from the soft science corpus show:

Example 1(soft science): "To what extent are ***heritage language acquisition (HLA) and second language acquisition*** (SLA) comparable in different linguistic domains?" (Albirini, Benmamoun, & Saadah, 2011, p.300)

Example 2 (soft science): "These descriptions of learner development facilitate a more comprehensive understanding of how GM emerges ***in second and foreign language learning*** and in turn, how learner syllabi need to adjust to incorporate these critical markers of language development" (Liardét, 2013, p.168).

Example 3(soft science): "The role of the first language has become an increasingly important subject of research ***in second and foreign language instruction***, with scholars investigating teachers' and students' codeswitching beliefs and practices in a range of settings".

The colligations of the collocation of 'language +second' can be written as follows:

Example 1: ***heritage language acquisition (HLA) and second language acquisition*** (SLA)

Adj+ n+ n+ connector (and) + adj + n +n

In example 1, the word 'second' is the adjective for the word of 'language' that the collocation of 'second language' plays the role of adjective and extra information for the noun of 'acquisition'. Furthermore, the adjective of 'heritage' is an extra explanation for the language as a noun that both of them (heritage language) give extra information for the noun of 'acquisition'.

Example 2: ***in second and foreign language learning***

Preposition + adj+ connector (and) + adj + n + n

In example 2, the word 'second' is placed after the preposition of 'in' and similar to the example 1, it is an adjective for the word 'language' that is accompanied with the connector and the second adjective (foreign) followed by the other noun that is 'learning'.

Example 3: ***in second and foreign language instruction***

Preposition + adj+ connector (and) + adj + n + n

In the example 3, the same rule was repeated and the collocation of 'second language' was distracted by the connector of 'and' and the other adjective that is the term of 'foreign'. In addition, the preposition of 'in' was placed before the term 'second'. In sum, based on the results, the dominant colligations in the left position of the collocation of 'second language' were prepositions, adjectives, possessive forms, verbs (to + v), conjunctions, and in some cases the symbol of 'dot' as the last point of the sentence. The dominant colligations in the right hand were low in numbers and most of the words followed the collocation of 'second language' were adverbs, connectors, (L2), and in most of the cases the symbol of 'dot'.

The same procedure was done for the hard science that based on the findings, the high frequent content word in hard science was 'patients. Figure 3 visualizes the top collocates of the noun 'patients' occurring immediately prior and subsequent to the node word in the hard sciences.

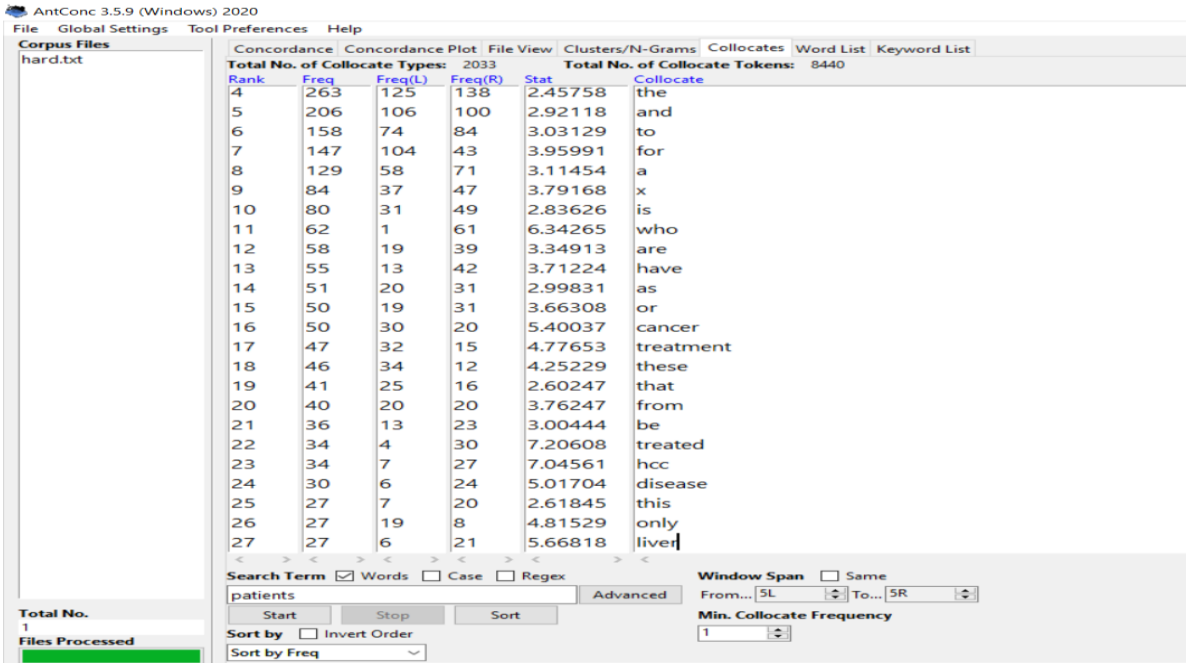

Figure 3

The top collocates of the noun 'patients' occurring immediately prior and subsequent to the node word in the hard sciences

Based on the analysis from the file view section of the Antconc software, the results showed that the collocate of ‘patients’ in hard science are observed in both left hand and right hand with different frequencies. As stated, to elaborate on the collocates of high frequent words in the corpora, one content word from each corpus were selected and analyzed. The high frequent collocation of the content word of ‘patients’ in hard science was ‘cancer’ that the examples were indicated in Figure 4.

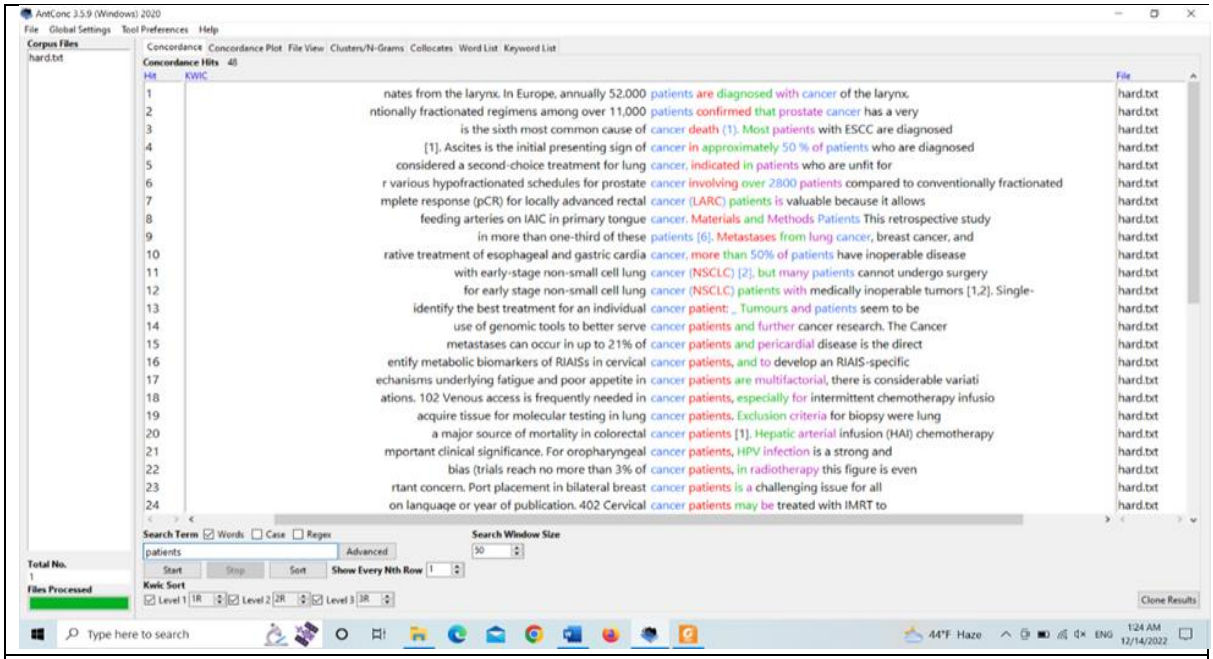

Figure 4

Instances of the collocate of 'patients + cancer' in hard science

As figure 4 visualizes, in hard science, the collocate of ' patients + cancer' are observed in both left hand and right hand, however, the weight of left hand (F=30) was higher than the right hand (F=20). It means that cancer can placed in the left hand of the content word of

'patients' as the following three instances from the hard science corpus show:

Example 1 (hard science): "In Europe, annually ***52, 000 patients are diagnosed with cancer of*** the larynx."

Example 2 (hard science): "…among over ***11,000 patients confirmed that prostate cancer*** has a very high sensitivity to dose per fraction".

Example 3 (hard science): "…and is the sixth most common cause ***of cancer death (1). Most patients with*** ESCC are diagnosed at an advanced stage."

The colligations of the collocation of ''patients + cancer' in hard science can be written as follows:

Example 1: ***52, 000 patients are diagnosed with cancer of***
Number+ n (patients)+ to be+ v+ proposition+ n (cancer) + preposition

In this example, the word 'patients' is preceded by number and followed by to be (are). Besides, the word 'cancer' centered by two prepositions and acted as the noun for the first preposition and as the extra information for the second one that was followed by the terms 'the larynx'.

Example 2: ***11,000 patients confirmed that prostate cancer***
Number + n (patients) + v+ conjunction (that)+ n+ n (cancer)

In example 2, the word 'patients' is placed after the number and followed by a verb plus conjunction in order to be connected the collocation of 'prostate cancer'. As it is axiomatic, the term 'cancer' acted as a noun and followed by the other noun that gave the extra information for 'cancer'.

Example 3: ***of cancer death. Most patients with***
Preposition +n (cancer) + n + dot+ adj+ n (patients) + preposition

In the example 3, the term 'cancer' was preceded by preposition and followed by a noun at the end of the sentence. However, the term 'patients' preceded by an adjective and followed by a preposition. In general, based on the results, the dominant colligations in the left position of the collocation of 'patients + cancer' in hard science were prepositions, numbers, nouns, connectors, and verb. The dominant colligations in the right hand were nouns, adverbs, connectors, to be, pronouns, conjunctions, comma, verb, and in some cases the symbol of 'dot'. Based on the findings, the dominant collocations in two corpora were grammatical collocations since the function words were the high frequent words in two sciences. Moreover, the high frequent colligations as the function words in soft science were 'the, and, of, in, to' similar to those of hard science. To be concluded, the authors in both disciplines used different grammatical forms before and after the collocations without significant differences. In other words, there was no substantial difference in two sciences in replacing grammatical forms in the left and right positions of collocations.

DISCUSSIONS

As stated, the main objective of the current study was to explore the rates of collocations and colligations used in the introduction sections of two corpora of soft and hard sciences based on Hoey's lexical priming model. Furthermore, exploring the high dominant collocations and colligations were the other aims. Finally, exploring the existence of any significant difference in terms of collections and colligations in two corpora was the last goal. The findings revealed that the frequency of collocations that employed in the introduction sections of soft and hard sciences were not the same and the researchers of soft science in designing their research articles' introduction sections employed more collocations in comparison to the researchers in hard science. Hence, based on the results of Chi-square data analysis, it can be claimed that there is a significant difference between two groups of authors in using colocations. The justification can be the familiarity of two groups of authors with academic writing as well as the use of collocations. The other reason may go to the high number of collocations in the introduction sections written by soft science authors. The results of this study provide strong assistance

to the theory of lexical priming put forward by Hoey (2005). A term is collectively filled with the contexts and co-texts where it is presented, as per this theory, and our interpretation of it involves the fact that it co-occurs in certain kinds of context with certain other words. In other words, a term is thus primed for its collocates and for the right locations or positions of those collocates. In technical terms rather than in general words, the power of co-occurrences of collocates was high. Knowledge of collocations is deemed a compulsory element of academic reading and writing skills that can be closely attributed to the job opportunities, academic success, economic well-being and public status of individuals (Selmistraitis, 2020).

However, in terms of colligations, both groups of authors used colligations with the same rates and similarities in the right and the left positions of the collocations. In the current study, the high frequent content words of each corpus were studied and analyzed in detail with the other content word to determine the dominant colligates. Based on the findings, the dominant collocations were grammatical collocations in two corpora since the function words were the high frequent words in two sciences. Moreover, the high frequent colligations as the function words in soft science were 'the, and, of, in, to' similar to those of hard science. As it is axiomatic there is no difference in two sciences in terms of high dominant colligations. In fact, the same pattern was repeated in two corpora. The results are in congruent with Paškevičiūtė's study (2020). It attempts to fill the gap on the comparison of popular scientific and academic discourses as well as hard and soft sciences through the use of collocations. The aim of the research was to analyze and compare collocations with the most frequent nouns and adjectives in hard and soft science disciplines in popular scientific and academic discourses in English in order to gain more insight into the differences and similarities between both discourses and hard and soft sciences. The analysis has revealed some similarities between academic and popular scientific discourses and hard and soft sciences. As for the differences in hard and soft sciences, some collocations with the most frequent nouns and adjectives could be regarded as discipline-specific as the texts in hard sciences seem to frequently mention health, system and environment-related aspects which could typically be found in science, technology and medicine disciplines, while the texts in the soft sciences focus more on educational, political, societal and minority-related issues, which might predominate in disciplines such as education, political science and social sciences.

The results are in congruent with a study conducted by Haghverdi, Biriya, and Alipour (2013). This study attempted to compare the colligations of the noun group in two specialized corpora of hard and soft sciences to find out whether they are realized similarly or differently in different disciplines. Having counted the occurrence frequencies of the collocation categories of the word, the researchers could obtain solid evidence to claim that the word group resorts to different colligational patterns across hard sciences and soft sciences. The findings of the present study also lend great support to the lexical priming theory put forth by Hoey (2005). According to this theory, a word becomes cumulatively loaded with the contexts and co-texts in which it is encountered, and our knowledge of it includes the fact that it co-occurs with certain other words in certain kindsof context. Therefore, a word is primed in the cont ext of its occurrence for its colligations. That is why the word group, for instance, employs a number of different colligational patterns in chemistry and physics texts compared to psychology and sociology texts.

Collocations and colligations are of prime importance for non-native speakers of English. As Demir (2017) stated one of the main obstacles, particularly for non-native writers (NNW), is indeterminate knowledge of word combinations. Through the acquisition of collocation, it may be possible for NNW to increase their lexical competence. Based on this claim, Demir (2017) attempted to investigate the use of English lexical collocations in the texts written by native writers of English (NW) and non-native writers of English (NNW), and to examine whether there are any statistically significant differences between NW and NNW in terms of employing collocations in their written productions. As it is clear, very similar to this study, Demir (2017)

compared two groups of authors in the use of collocations. The corpora for his study consisted of 40 research articles (RAs) published in leading journals in ELT, 20 of which belong to native speakers of English while the rest to non-natives. Based on the findings, Turkish authors used noun + verb collocations more than native authors. Meanwhile, the results put forth that native writers have a tendency of using adjective + noun collocations.

The results are in congruent with a study that recently has been done by Phoocharoensil (2020), which like this study, the researcher conducted a corpus-based study to examine genres and collocation patterns in which the three synonyms 'consequence', 'result', and 'outcome' usually occur. The difference of this study with Phoocharoensil's study (2020) is the fact that we compared two academic genres in the written format, however, Phoocharoensil (2020) analyzed three mentioned synonyms. The results showed that of all the eight genres currently available in COCA, the three synonyms appear with the highest frequency in academic texts, whereas frequencies are lowest in informal genres. To sum up, the findings of the current study will contribute to education in argumentative essays, that are a popular type of pedagogical writing allocated to undergraduate and language-focused programs across disciplines.

CONCLUSION

This study aimed at investigating the rates of collocations and colligations used in the introduction sections of research articles in soft and hard sciences based on Hoey's lexical priming model. Furthermore, investigating the high dominant collocations and colligations were the other goals of this study. Also, finding out any significant difference in terms of collections and colligations in two corpora was the other end. The results showed that in the introduction sections of research articles written by soft and hard sciences' researchers, although in terms of colligations, the same routine were repeated in two corpora, actually there was a significant difference between two groups of authors in the frequency of collocations. The researchers in soft science used more collocations than the authors in hard science. Finally, the results obtained showed that the dominant collocations were grammatical collocations in two corpora since the function words were the high frequent words in two majors. Moreover, the high frequent colligations as the function words in TEFL were 'the, and, of, in, to'. Also, as the results indicated, the high frequent colligations of TS as the function words were 'the, and, of, in, and'. It means that there is no difference in two majors in terms of high dominant colligations. In fact, the same pattern was repeated in two corpora.

We may draw a variety of conclusions taking into account the results of the distinctions made between how terms are understood in soft and hard disciplines with respect to the collocates. Firstly, it is shown that a phrase, such as 'language' or 'patients', appears to combine with some distinct and specific terms in general in soft and hard disciplines, and some distinct and special functions and content. Second, although they were overlooked in the current study, the frequencies of function words are high in two genres despite the content words that are not comparable in frequencies throughout two genres. In addition, it is worthwhile saying that since the data for the current research was gathered from ISI research articles as notable ones, hence reporting the frequencies of different categories of lexical priming such as collocations can have pedagogical implications for novice authors in both hard and soft science to pay attention to the high frequent categories and types and use them as models in their writings as a pattern. Given that the major problem in learning may lay in collocations, English instructors are advised to focus their efforts on assisting L2 students with word collocation. Instructors could use corpus-based research like the current one to enlighten students about the distinctions in collocations in terms of the majors they are studying to help them grasp the different uses of words, improve their own vocabulary, and therefore raise L2 knowledge. Also, unawareness of the nature of collocations, in terms of general and specific collocations related to each science, perhaps compensate for the abuse, overuses, and underuses of English collocations to a significant extent, hence studies

like the present one can assist learners and writers of research articles to scrutinize on the nature of collocations and colligations. The same study can be conducted with soft and high science paying attention to the other sections of research articles. Moreover, the culture was ignored in the current study that the interested researchers can scrutinize on this issue. Finally, further research could also look at the use of collocations and colligations in other genres such as dissertations and academic speeches and conferences to find out what type and functions are dominantly employed by the users.

References

Almela-Sánchez, M., & Cantos-Gómez, P. (2019). The Grammatical Environment of Intensifier-Noun Collocations: Insights from Lexical Priming Theory. In*International Conference on Computational and Corpus-Based Phraseology* (pp. 1-14). Springer, Cham.

Anthony, L. (2011). Antconc (Version 3.2. 4w) [Computer Software]. *Tokyo: Waseda University, [cit. 31.5. 2015]. Dostupné na: http://www. antlab. sci. waseda. ac. jp/antconc_index. html*.

Barlow, M. (2016). *Wordskew*: Linking corpus data and discourse structure. *International Journal of Corpus Linguistics*, *21*(1), 105–115.

Dahunsi, T. N., & Ewata, T. O. (2022). An exploration of the structural and colligational characteristics of lexical bundles in L1–L2 corpora for English language teaching. *Language Teaching Research*, 13621688211066572.

Gu, B. J. (2017). Corpus-based study of two synonyms: Obtain and gain. *Sino-US English Teaching*, *8*, 511-522.

Halliday, M. A. K. (1966). Lexis as a linguistic level' in C. Bazell, J. Catford, M. A. K. Halliday, and R. Robins (eds): *In Memory of J.R. Firth*. London: Longman, pp. 148–62.

Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hoey, M. (2003). Why grammar is beyond belief. In J.-P. van Noppen, C. Den Tandt, & I. Tudor (Eds.), *Beyond: New perspectives in language, literature and ELT* (*Belgian journal of English language and literature, new series, 1,* pp. 183–96). Ghent, Belgium: Academia Press.

Hoey, M. (2004). The textual priming of lexis. In S. Bernadini, G. Aston, & D. Stewart (Eds.), *Corpora and language learners* (pp. 21–41). Amsterdam, Netherlands: John Benjamins.

Hoey, M. (2005). *Lexical priming: A new theory of words and Language*. London: Routledge.

Hoey, M. (2012). *Lexical priming: A new theory of words and language*. Routledge.

Liardét, C. L. (2013). An exploration of Chinese EFL learner's deployment of grammatical metaphor: Learning to make academically valued meanings. *Journal of Second Language Writing*, *22*(2), 161-178.

Novakova, I., & Siepmann, D. (2020). Literary style, corpus stylistic, and lexico-grammatical narrative patterns: Toward the concept of literary motifs. In *Phraseology and Style in Subgenres of the Novel* (pp. 1-15). Palgrave Macmillan, Cham.

Pace-Sigge, M. (2018). *Spreading activation, lexical priming and the semantic web: early psycholinguistic theories, corpus linguistics and AI applications*. Springer.

Patterson, K. J. (2016). The analysis of metaphor: to what extent can the theory of lexical priming help our understanding of metaphor usage and comprehension? *Journal of psycholinguistic research*, *45*(2), 237-258.

Pecorari, D. (2006). Visible and occluded citation features in postgraduate second-language writing. *English for specific purposes*, *25*(1), 4-29.

Römer, U. (2022). Applied corpus linguistics for language acquisition, pedagogy, and beyond. *Language Teaching*, *55*(2), 233-244.

Subtirelu, N. C., & Baker, P. (2017). Corpus-based approaches. In *The Routledge handbook of critical discourse studies* (pp. 106-119). Routledge.

Thompson, P. (2014). Exploring Hoey's notion of text colligation in a corpus of student

writing. In *Diachrony and Synchrony in English Corpus Linguistics* (pp. 347-371). Peter Lang.

Biodata

Leila ghaemi is a PHD candidate in Tabriz Islamic Azad university in TEFL . she has been researching in different areas of teaching and he is interested in discourse analysis and corpus-based studies.
Email: leili_gha2004@yahoo.com

Nesa Nabifar is an assistant professor at Islamic Azad University, Tabriz Branch, Iran. She holds PH.D in linguistics and applied linguistics. Her areas of research are second language acquisition, first language acquisition and discourse analysis. She teaches linguistics, academic writing and language teaching courses at graduate and post graduate levels. She has published a lot of papers and has attended different international and national conferences.
Email: n.nabifar@iaut.ac.ir

Davud kuhi is academic member of the department of English language in Islamic Azad universirty. He researched on different properties of academic discourses for approximately two decades and has published many articles in prestigious national and international academic journals.
Email: davudkuhi@yahoo.com

Creative Commons License: *This work is licensed under a Creative Commons Attribution-Non Commercial-ShareAlike International License*