IRJABS

Textual Positions of Lexical Bundles across Newspaper Genres

Siamak Ansari[1*], Arezoo Molavi[1]

1. Department of English Language, Omidiyeh Branch, Islamic Azad University (IAU), Omidiyeh, Iran

Corresponding author email: ansari@iauo.ac.ir

ABSTRACT: According to Hoey (2005), words or their combinations may be primed to occur or avoid occurring at the beginning or end of independently recognized discourse units, like the sentence and the paragraph; these are their textual colligations. This study attempted to investigate the textual positions of two English lexical bundles across two newspaper corpora, each including approximately one million words. The genres considered were economic and political news report, recently downloaded from internationally famous newspapers. To this end, Hoey and O'Donnell's (2008) dichotomy of textual positions was employed for all the analyses. According to this framework, textual positions comprise: first sentence of text, first sentence of a paragraph that is not the first in text, and any sentence that does not begin a paragraph or a text. Furthermore, based on Hyland's (2008) taxonomy, two lexical bundles which frequently occurred in the two corpora were selected. In addition, the Wordsmith Tools software (Scott, 2010) was used to conduct the comparisons. The frequency and percentage of each bundle was calculated; then chi-square was applied and the results revealed significant generic differences in the use of the bundles with regard to their textual positions.
Key words: Colligation, textual positions, lexical bundles, genre, newspaper reports

INTRODUCTION

The notion of priming is based on the assumption that "the mind has a mental concordance of every word it has encountered, a concordance that has been richly glossed for social, physical, discoursal, generic, and interpersonal context" (Hoey, 2005, p. 11). In fact, it is the argument of this theory that priming is the driving force behind language use, language structure and language change. Hoey's 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, 2004, p. 386).

According to Sinclair (1991), speakers perform with an idiom principle, interconnecting context-bound lexical choices and only considering the grammar as a last resort. According to this view, while we are speaking, we identify the most associated words and phrases and, subsequently, depending on our textual experience, affect our usage. Thus, speakers associate particular words and phrases with specific textual positions, and so to be a good writer, we need something more than deciding on grammatical patterns and on collocations. That is why some texts are more natural and others more unnatural and odd to us (Hoey & O'Donnell, 2008).

Hoey (2005) defines textual colligation as 'every lexical item (or combination of lexical items) is capable of being primed (positively or negatively) to occur at the beginning or end of an independently recognized "chunk" of text'. When we come upon language in speech or writing, we are aware of the contexts and co-texts in which we encounter it. The point is that just as we are aware that words are typically used as part of Subjects or Adjuncts, so we are also aware of their textual position. Therefore, for instance, our awareness includes the knowledge that (as we have already seen) certain words tend to come at the beginning of a sentence, while others tend to come at the end of a sentence (pp. 129-130).

Lexical bundles, as a particular and relatively recent category of word combinations with a possibly formulaic status (see Biber & Barbieri, 2007), were coined and defined by Biber, Johansson, Leech, Conrad, and Finegan (1999). They defined lexical bundles as "recurrent expressions, regardless of their idiomaticity and regardless of their structural status" (p. 990). They refer to frequency as the most salient and defining characteristic of bundles; for example, the word combination on the other hand which is considered as a bundle, must at least

occur tentimes in a corpus made of one million words and must be realized in at least five different textsto guard against idiosyncratic or repetitive uses.

To produce texts successfully, writers should be aware of the context and the readers whoform part of that context. The importance of genre, as Bianchi and Pazzaglia (2007) believe, liesin the fact that genre conventions and accepted rhetorical practices may be valuable resources forwriters in general and non-native speakers in particular. Besides the rhetorical structures of textswith which the majority of genre studies deal, successful writers have to be aware of the waysthrough which lexical items are put to use in any specific domain or genre. Otherwise, theirproducts would seem unnatural and weird. Textual positions are very significant in this regard torender texts palatable and natural in a given genre.

Agreements of academic writing are variously exploited in different disciplines (Hyland, 2005 & 2009). Therefore, it is essential for writers to have a clear idea of textual positions of lexical bundles in English which is generally the shared language of research. Moreover, writers would take advantage of knowing how lexical bundles are used in different disciplines.

Thus, the present study intends to investigate the textual positions of English lexical items acrosstwonewspaper corpora considering economic and political news reports to explore how significantly the textual positions of lexical bundles across various genres differ.

Lexical priming theory is an attempt to illustrate how naturalness is achieved and how an explanation of what is natural might impinge on explanations of what is possible. In this regard, a key factor in naturalness is collocation(Hoey, 2005, p. 2).

Lexical priming not only accounts for collocation, but it can also describe everything that we know about a word. The notion of priming is based on the assumption that "the mind has a mental concordance of every word it has encountered, a concordance that has been richly glossed for social, physical, discoursal, generic, and interpersonal context" (Hoey, 2005, p. 11). In fact, it is the argument of this theory that priming is the driving force behind language use, language structure and language change.

According to Hoey, priming accounts for the following features:First, every word is primed to occur with particular other words; these are its collocates.Second, every word is primed to occur with particular semantic sets; these are its semantic associations. Third, every word is primed to occur in (or avoid) certain grammatical positions, and to occur in (or avoid) certain grammatical functions; these are its colligations (Hoey, 2005, p. 13).

The notion of colligation has its origin in Firth, who introduced it, "the statement of meaning at the grammatical level is in terms of word and sentence classes or of similar categories and of the inter-relation of those categories in colligation […]." (Firth, 1957, as cited in Hoey, 2005, p.13).Although this notion has long been recognized in corpus linguistics, surprisingly little substantial work exists explicating and/or applying it.

So far colligation was defined as the grammatical associations a word or word sequence is primed to favour or avoid, but it is significant that Halliday (1959) also formulates the colligational relationship in terms of sentential position. This is an important extension. It 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.

There are textual implications to lexical priming theory, namely that certain words and phrases have associations with the beginnings of texts and paragraphs. Lexical items are primed to appear in text-initial or paragraph-initial position (Hoey,O'Donne, Mahlberg& Scott 2007). They related these lexical and corpus-driven facts to current textual descriptions of news stories that might provide explanations for the positive priming of relevant lexis. They examined words in terms of their positions, text-initial sentences (TISC), paragraph-initial sentences (not in TISC) (PISC) and non-initial sentences (NISC), and noticed items that were 'key' in TISC (Text-Initial Sentence Corpus) when compared with NISC (Non-initial Sentence Corpus), for example, yesterday and last night;announced, suffered, stressed, and according to a;plans and report. They found out three ways a key TISC word might differ from its NISC counterparts. First, it might simply differ in frequency. Second, it might differ proportionally in use, or certain usages might be absent from one or the other corpus, but the difference is a matter of degree. Third, it might differ from its NISC counterparts.that is, it might differ markedly in its use in TISC from NISC.

Yamasaki (2008) investigated the association of particular collocations and colligations with discourse functions of unspecific anaphoric nouns. He used the British National Corpus (BNC) and the concordancing program SARA to identify the behavioral patterns amonglabels in relation to other lexical items and grammatical categories. He also compared the characteristic use of labels in spoken discourse with their use in typical writing. The corpus-based analyses revealed that typical collocational and colligational patterns and functions of lables differ between writing and speech. The resultsshowed that collocations and colligations associated with a word may be avoided by another as they have distinct discourse functions in specific contexts.

Gabrielatos (2007) examined the case for treating if-conditionals as strong attractors of modality. If-conditionals, and if-constructions in general, can be seen as modal colligations, and this is a strong claim to be examined. The tree notions used in this study were: grammatical constructions, colligation, and semantic preference.The corpora used in this study were: a sample of 1,000if-constructions from the written BNC, the written BNC Sampler, FLOB, all the if-sentencesfrom the written BNC, and the non-conditional if-sentences from the sample. Further testsinvolve frequency comparisons of specific modal words between the manually annotatedsample and the annotated versions of BNC, BNC Sampler and FLOB, as well as a collocational analysis of if in the written BNC.The results of this study supported the claim that if-conditionals, seen as a group, carry a significantly higher modal load than average.

Hoey (2004) argued for a perspective on text-linguistic, rooted in lexical item that makes no distinction between the description of the text and the description of its component lexis. He tried to demonstrate that the properties of text can all be dealt with the concept of textual colligation. He argued that lexis is primed for textual use, which the choice of a lexical item is concurrently the choice of its priming. He discussed that any lexical choice positively or negatively has a major effect on features such as cohesion, theme choice and textual divisionsand that corpus investigation can shed light on the nature of the lexical choices made. Hoey also argued that the lexical choices interlock, creating what Hoey termed "colligational prosody".

Hunston, Francis and Manning (1997) concluded that there are correlations between grammatical patterns and lexical meaning. That means a word could have a specific meaning when it co-occurs with a certain word. This hypothesis is followed by Hoey (2000), who mentioned that some meanings of the same word have their own grammatical patterns, which is called "colligation".He showed how corpus analysis revealed previously unnoted types of regularity and argued these rules should be taught to learners. In his paper, he attempted to account word combinations for students' reconstructions of a short text whose paragraphs had been jumbled. To explain why certain paragraphs were interpreted as possibly or not possibly text-initial, he extended the notion of colligation to cover not merely the relationship of a word to the structure of the clause and sentence, but also the paragraph and the entire text, as reflected in its tendency to occur in particular textual positions. He arguedthere must be a "hidden colligational signaling" of the text structure, unknown to us, but of clear pedagogical relevance for the teaching of reading and writing (Hoey 2000, cited in Meg Lu et al., 2011).

Tono (2007) aimed to reveal some overuse and underuse patterns of colligations and collocations and to identify possible causes of those uses to either developmental patternsbetween L2 learner corpora and L2 learners' mother tongue corpora (in this case, Japanese). This study shed more light on the nature of inter-language development and proposed a new approach toward the study of learner language.

In one study,Rafiee and Tavakoli (2011)explored Hoey's (2005) notion of textual colligation, by examining the textual locations and environments of a variety of single and multiword items in a corpus of undergraduate assignments, derived from the British Academic Written English (BAWE) corpus.They identified where the items occur in the text and where the items occurin a paragraph. They also examined the textual environments in which these items occur, both at a sentence level and at a broader textual level. The BAWE corpus contained a large number of texts taken from different disciplines and representing a range of genres. Furthermore, they discussed the evidence for disciplinary variation, and whether words or phrases are primed with different textual colligations, depending on discipline and genre. Rafiee and Tavakoliargued that Hoey's (2005) theory helps to reunite corpus-derived observations about language with concepts of rhetoric, and that textual colligation is bound up with experiences and understandings of how texts are organized.

There have been some studies that have developed functional classification of lexical bundles (e.g., Cortes 2001, 2002, 2004; Biber, Conrad, and Cortes, 2003; Biber et al, 2004; Biber and Barbieri, 2007; Hyland, 2008). Such studies have proved that these word clusters can serve such a wide range of discursive functions as organization of discourse, expression of stance, and reference to textual or external entities.

Some studies explored on the use of lexical bundles in master theses, doctoral dissertations, and research articles in the discipline of applied linguistics (Jalali, EslamiRasekh&TavangarRizi, 2008, 2009; Jalali&Ghayoomi, 2010). Moreover, (Juknevicien, 2009; Ping, 2009; Rafiee&Tavakoli, 2011), compared the use of lexical bundles in the writing of native and non-native corpora to see whether the non-natives use of lexical bundles match the way natives use them.

Reviewing the related literature reveals an evident gap in the previous studies. Although there are many studies on lexical bundles, however studies on these bundles in terms of textual positions are rather rare. Also,several works have been made to decide about the relationship between lexis, grammar andcollocation, while, to the best of our knowledge, very few, if any of them, have been oriented toinvestigate the relationship between textual positions and lexical bundles in newspaper reports.It seems that words and lexical bundles (3-word, 4-word,

and 5-word) may be realizeddifferently across different disciplines in terms of their textual positions (Hoey 2005). This studyaims to find out how textual positions might vary in newspapers of two different genres.

By analyzing an adequate number of newspaper reports, the present study compares and contrasts the textual positions of lexical items in two different genres –political/ economical– of English native authors. This study serves to answer the following question:

Do the textual positions of lexical bundles in the newspaper reports of different genres differsignificantly?

METHOD

Instrument

The two frameworks used to analysis the textual positions are described here. One is theHoey and O'Donnell's (2008) dichotomy of textual positions of lexical bundles as presented below. According to Hoey and O'Donnel (2008), texts normally consist of a series of paragraphs, which in turn consist of sentences. They lable these sentences T, P, or N depending on their location in a paragraph and text. A text-initial sentence (T) is the first sentence of the text.

Paragraph initial sentences (P) are the first sentence of any other paragraphs. Finally, non-initial sentences (N) are sentences that are not the first sentence of the paragraph in which they occur.

Each sentence can be divided into two halves, which they number 1 and 2. Thus, for example, T1 refers to the first 50% of a text-initial sentence and N2 refers to the second 50% of a non-initial sentence.

Table 1. Text, Paragraph, Sentence Scheme

	First half of sentence	Second half of sentence
First sentence of text	T1	T2
First sentence of paragraph that is not the first in text	P1	P2
Any sentence that doesn't begin a paragraph or text	N1	N2

Some examples from different disciplines are included to clarify these positions:

AL: ….. . Accordingly, it is necessary to consider the texts of these ads **in order to**determine the images of assisted reproduction that they construct. (N2 position)

Physics: … .On the other hand, when the PEC shell is put inside the cloak, thebeam is guided around the shell and leaves the cloak without significantperturbation. (N1 position)

AL: In order to compare the reading comprehension of the male subjects on thethree different text types, a one-way ANOVA was conducted, the results of which aredisplayed in Table 2. (P1 position)

AL: The collected data in this study was coded on the basis of the coding schemedeveloped by CCSARP with some modification (Blum-Kulka and House,1989). (P2 position)

The other instrument was Hyland's (2008) taxonomy of the most frequent 3-word, 4-wordand 5-word lexical bundles, illustrated in the following table.

Table 2.Hyland's (2008) classification of the most frequent 3, 4 and 5-word lexicalbundles in academic corpora

3-word	4-word	5-word
in order to	on the other hand	on the other hand the
in terms of	at the same time	at the end of the
one of the	in the case of	it should be noted that
the use of	the end of the	it can be seen that
as well as	as well as the	due to the fact that
the number of	at the end of the	at the beginning of the
due to the	in terms of the	may be due to the
on the other	on the basis of	it was found that the
based on the	in the present study	to the fact that the
the other hand	is one of the	there are a number of
in this study	in the form of	in the case of the
a number of	the nature of the	as a result of the
the fact that	the result of the	at the same time the
most of the	the fact that the	is one of the most
there is a	as a result of	it is possible that the
according to the	in relation to the	one of the most important
the present study	at the beginning of	play an important role in
part of the	with respect to the	can be seen as a
the end of	the other hand the	the results of this study
the relationship between	the relationship between	the from the point of view
in the following	in the context of	the point of view of
the role of	can be used to	it can be observed that
some of the	to the fact that	this may be due to
as a result	as shown in figure	an important role in the
it can be	it was found that	in the form of a

According to Hyland (2008), this framework is on the basis of frequency and the first 10 bundles were the most frequent ones. In order to analyze the textual positions of lexical bundles in the two disciplines , we chose the most frequent of each group (3-, and 4-word) of lexical bundles to find out whether the most frequent position of each bundle is T, P, or N. The two most frequent 3-word lexical bundles were accordingto theandone of the.

Moreover, the fifth version of the WordSmith Tools software (Scott, 2010) which is an integrated program for looking at how words behave in texts was employed in this study. This software allows us to find out how words are used in texts and also to find the position of words in each corpus.

RESULTS AND DISCUSSION

This section reports on the data obtained through the study as well as the statistical measures to find out whether significant differences existed between the textual positions of the three-word lexical bundles 'according to the' and 'one of the' in the context of economic vs. political news reports. The raw frequencies and percentages of the lexical bundles in the disciplines with regard to their occurrences in T1, T2, P1, P2, N1 and N2 positions are summarized in the tables below.

Table 3. Results of analyzing textual positions for 'according to the' in the two different corpora

Position Corpora	T1	T2	P1	P2	N1	N2	Total	Sig.
Economic	0 (0%)	0 (0%)	20 (3.38%)	30 (5.08%)	150 (25.42%)	390 (66.12%)	590	
Political	30 (5.88%)	10 (1.96%)	90 (17.64%)	110 (21.56%)	70 (13.72%)	200 (39.24%)	510	0.0

Table 3 reveals that the 3-word lexical bundle 'according to the' was more frequently used by economic reporters, 530 times in this corpus. We can see that, in both genres, this lexical bundle was more likely to occur in the second half of any sentence that does not begin a paragraph or text (N2 position). Another fact is that this bundle, in both genres, did not occur in T1 position at all. Table 1 indicates that the significance value for the two genres was 0.0 which is smaller than 0.05. Chi-square analysis revealed a significant difference in the use of this bundle in different textual positions between the two genres.

The following table 4 represents the findings for the other lexical bundle 'one of the' across the above corpora.

As illustrated in table 4, the 3-word lexical bundle 'one of the' was recorded 590 times in the economical context, and 510 times in the political one. Similar to the case for 'according to the', this lexical bundle occurred more in the second half of any sentence that does not begin a paragraph or text (N2 position) in both genres. However, this time there is a remarkable difference in the frequencies, as one can see 'one of the' was repeated 390 times in N2 position and only 150 times in N1for the economic genre in comparison to political one which were 200 times in N2 position and 70 times in N1. Another point is that this bundle did not occur in T1 position in the political genre as the previous result for 'according to the' represented, but it did occur in T2 position.This bundle was more utilized in the second half of the first sentence of a paragraph that is not the first in the text (P2 position) in Economic corpora was 30 for economic compared to the other corpus which was only 110 for the political.

Table 4 denotes the significance value for the two genres was 0.0 which is smaller than 0.05. Chi-square analysis revealed a significant difference in the use of this bundle in different textual positions of the two genres.

CONCLUSION

Bearing in mind the deficiencies of works on textual positions of lexical bundles leads us to the importance that the present study has in fostering our knowledge of textual positions of lexical bundles in different genres. It is worth mentioning that every lexical item (or combination of lexical items) is capable of being primed to occur at the beginning or end of an independently recognized "chunk" of text.

The results showed that lexical bundles are primed differently in different disciplines with regard to their positions in text. In this study, the Economic and Political newspaper reportswere selected as the target genres. It was revealed that the total number of lexicalbundles used in different disciplines and also the uses of specific lexical bundles in differenttextual positions were not similar. This may be due to the fact that each discipline represents adifferent kind of research and methodology.

Table 4. Results of analyzing textual positions for 'one of the' in the two different corpora

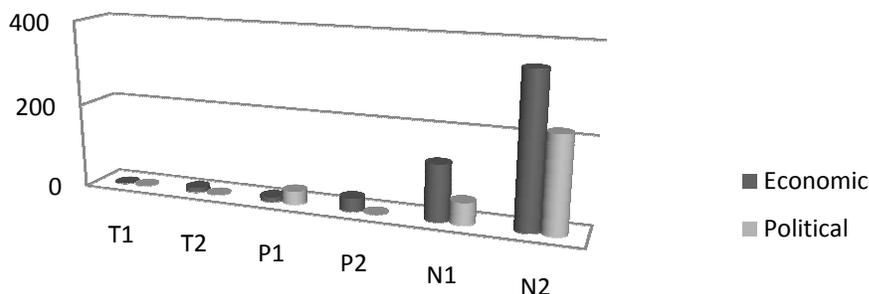| Position Corpora	T1	T2	P1	P2	N1	N2	Total	Sig.
Economic	0 (0%)	10 (1.88%)	10 (1.88%)	30 (5.66%)	130 (24.52%)	350 (66.06%)	530	0.0
Political	0 (0%)	0 (0%)	30 (10%)	0 (0%)	50 (16.66%)	220 (73.34%)	300	

Figure 1. Textual positions of *'according to the'* in economic vs. political corpora

By this is meant the total number of 'according to the', forinstance, in Economic corpora was more than in the other genre. Also, the authors of differentacademic disciplines use specific norms of writing which belong to that specific discipline andonly the members of that discourse community are familiar with those conventions. Therefore, itcan be said that lexical bundles are primed differently in terms of their textual positions bymembers of each discourse community.

The results of this study showed that the frequency of using lexical bundles in different textual positions varies both in different disciplines and that the writers' use of lexical bundles in different positions in RAs might vary according to the disciplines they are writing in.

However, as Hoey (2005) puts it, 'priming is what happens to the individual and is the direct result of a set of unique, personal, unrepeatable and humanly charges. Therefore, large-scaled investigations into more conditioned experiments need to be conducted. The writersof reports or any piece of writings which are to be investigated should come from different nationalities regarding their being native/ nonnative speakers, with certain degrees of education, etc.

REFERENCES

Bianchi F, Pazzaglia R.2007. Student writing of research articles in a foreign language: Metacognition and corpora. In R., Facchinetti (Ed.), Corpus linguistics: 25 years on. (pp. 259-287). New York: Rodop.

Biber D, Barbieri F.2007. Lexical bundles in university spoken and written registers. English for Specific Purposes, 26, 263-286.

Biber D, Conrad S, Cortes V.2003. Lexical bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson& T. McEnery (Eds.), Corpus linguistics by the lune: A festschrift for Geoffrey Leech, 71-92. Frankfurt: Peter Lang.

Biber D, Conrad S, Cortes V.2004. If you look at...: Lexical bundles in university teaching and textbooks. Applied Linguistics, 25, 371–405.

Biber D, Johansson S, Leech G, Conrad S, Finegan E.1999. Longman grammar of spoken and written English. Harlow: Pearson.

Cortes V.2001. Lexical bundles in context: A new taxonomy. Unpublished Manuscript, Northern Arizona University, Arizona.

Cortes V.2002. Lexical bundles in published and academic student writing in history and biology. Doctoral dissertation, Northern Arizona University.

Cortes V.2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. English for Specific Purposes, 23, 397–423.

Halliday MAK.1959. The language of the Chinese: Secret history of the Mongols. Oxford: Blackwell.

Hoey M, O'Donnell MB.2008.Lexicography, grammar, and textual position. International Journal of Lexicography, 21, 293-309.

Hoey M.2000. A world beyond collocation: new perspectives on vocabulary teaching. In M. Lewis (Ed.) Teaching collocation.Hove: LTP.

Hoey M.2004.Lexical priming and the properties of text. In A. Partington, J. Morley & L. Haarman (Eds.), Corpora and discourse. (pp. 385–412). Bern: Peter Lang.

Hoey M.2005. Lexical priming: A new theory of words and language. London: Routledge.

Hoey M, O'Donnell MB, Mahlberg M, Scott M.2007. Exploring the textual positions and functions of lexical items in hard news stories. Corpus Linguistics.University of Liverpool.

Hunston S, Francis G, Manning E. 1997. Grammar and vocabulary: Showing the connections. ELT Journal, 51(3), 208-216.

Hyland K.2005. Metadiscourse. London: Continuum.

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

Hyland K.2009. Writing in the disciplines: Research evidence for specificity. Taiwan International ESP Journal, 1(1), 5-22.

Jalali H, Ghayoomi S.2010. A comparative qualitative study of lexical bundles in three academic genres of applied linguistics. MJAL, 2(4), 323-333.

Jalali H, EslamiRasekh A, TavangarRizi M.2008. Lexical bundles and intradisciplinary variation: the case of applied linguistics. Iranian Journal of Language Studies (IJLS), 2(4), 447-484.

Jalali H, EslamiRasekh A, TavangarRizi M.2009. Anticipatory 'it' lexical bundles: A comparative study of student and Published writing in applied linguistics. Iranian Journal of Language Studies (IJLS), 3(2), 177-194.

Juknevicien R.2009. Lexical Bundles in learner language: Lithuanian learners vs. native speakers. KALBOTYRA, 3, 61-72.

Lu M, Hui Lin C, Yen T, Ku T, Min Tsai C.2011. Assemble Collocation and Colligation in Chinese Writing Web Tools for New Immigrants in Taiwan. US-China Education Review, (8)5, 659-665.

Ping P.2009. A study on the use of four-word lexical bundles in argumentative essays by Chinese English majors: A comparative study based on WECCL and LOCNESS. CELEA Journal, 32(3).

Rafiee M, Tavakoli M.2011. Structural analysis of lexical bundles across two types of English news papers edited by native and non-native speakers. MJAL, (3), pp.136-155.

Sinclair J.1991. Corpus, concordance, collocation. Oxford: Oxford University Press.

Tono Y.2007. Japanese EFL learners' colligation/collocation patterns: Multiple comparison

Yamasaki N.2008. Collocations and colligations associated with discourse functions of unspecific anaphoric nouns. International Journal of Corpus Linguistics (pp. 75-98). John Benjamins Publishing Company.