



STUDY OF COLLIGATION: A COMPARISON BETWEEN NATIVE LEARNERS AND NON-NATIVE PAKISTANI LEARNERS

Amina Mumtaz

aminamumtazskt@outlook.com

Lecturer (visiting) Baba Guru Nanak University, Nankana Sahib

Abstract

When two lexical items are co-occur through grammatical association that is called colligation. Colligations can enhance the potential by binding of the words with different structures by lexical items. The purpose of this research study has drawn the comparison in between the native learners and nonnative Pakistani learners of English. The corpus has extracted from the ICNALE written essays. In ICNALE written essays, we have extracted the 400 files from native learners and 400 files from nonnative Pakistani learners. We have drawn the comparison. All these corpus files have tagged with the TagAnt, and identified the problem to use the colligation in the nonnative Pakistani learners. In this comparative research study, we have identified the misplacement of lexical items and erroneous of colligation in the Pakistani learners. The reason of this problem is the attitude of syllabus designers and influence of L1 background of Pakistani learners.

Keyword: TagAnt; corpus; ICNALE (International Corpus Network of Asian Learners of English); Colligation

1. Introduction

1.1 Background of the study

In Pakistan and other developing countries, it has used as the political and official language and considered as the language of development. English is the language which has more nonnative users than the native users. According to the statistical believe has four time more strength of nonnative users than the native users. In Pakistan, English has not a status of native language but it has been nativised (Hassan, & Mahmood, 2014). English has been codified and it is considered as the separate variety of language in Pakistan.

According to the Kachru's model, English has travelled around three circles of development at the globe and it has developed the three layers of the circle in his model. There are inner layer, outer layer and expanding layer. The inner layer has covered the native countries while second layer has covered that countries which are usually placed English as the second language and English is considered as the separate variety in all these countries. Pakistan has placed in the outer layer of the circle. Similarly expanding layer shows that countries which are considered English as the foreign language. In the Kachru model of English, Pakistan has placed as the outer layer as a second language country due to the high status of English in Pakistan.

English is used as the medium of instructions so it is considered as the second language in Pakistan (Rehman, 2020). According to national education policy of 2009, English has considered as the medium of instructions not only taught as the subject of foreign language (Channa, 2017).

According to the present language policy, it is not clear as the English is the separate variety. The policy makers and syllabus designers are not clear about their variety of English and they have not appropriate knowledge about the variety of English which they have used and why they have used this variety (Hassan, & Mahmood, 2014). English is used as the medium of

instructions and it has proved as the status of English as second in Pakistan. (Celce - Murcia&McIntosh, 1991)

Nonnative users of English have the foreignness to their competency of language and they have not awarded to the knowledge of lexical items which are co-occur together and termed as collocation (Shah, 2020).

Teacher and syllabus designers have used the traditional approaches and methods which are commonly suggested the grammar translation method which has not focused on the fluency of the language(Shah,2020) many studies have proved that English is considered as the separate variety which countries are usually in the place of second layer of Kachru’s model. In Pakistan there is still work to be done to investigate the effect of modern ELT method and approaches on the lexical items which have the natural tendency to occur one after another termed as collocation.

Similarly, Firth has introduced the new term of colligation which is nearly close to collocation and it has natural tendency to tie each other. The term colligation is the grammatical association between two words. (Nordquist, 1996)

Table 1.1 English Colligations

English colligation	Example
When two words are co-occur and grammatical tied to each other termed as colligation. Colligation is on the syntactic level.(Romer, 2020)	Go to school.

In this study, we have explored the knowledge of colligation (which is not a general term like a collocation. The nonnative Pakistani learners have familiarity about the knowledge of colligation. Pakistani learners have not familiarity about the knowledge of colligation. Pakistani learners have the different lexical behavior while using the knowledge of colligation. In Pakistan, English has the strong influence on the learning and teaching but the nonnative learners of Pakistan have still unaware to the knowledge of colligation. They have used the different pattern of colligation but they are not familiar to the lexical behavior in the pattern of colligation which has strong association to each other.

Colligation plays the important role in what it means to know language .colligation has the three distinct features of distributional attraction between the linguistics items (a)the relationship between the lexical items and the grammatical context For Example (consequence +BE+ subordinate clause) (b)The particular grammatical items and the lexical items have the relationship in which items have used (c) the lexical items have the relationship between the syntactic function and the particular grammatical item which it is used (Hoey,2005).

In this research study, we have explored the comparative study of colligation in between the native learners and nonnative learners of Pakistani English. We have drawn different patterns of colligation and we are able to know the use of colligation in between the native learners and nonnative learners of Pakistani English. The native learners have explored the knowledge of colligation in different way while on the other hand nonnative learners of Pakistan have not used the colligation like the native learners of English. So this research study has drawn to use the knowledge of colligation in between native learners of English and nonnative learners of Pakistan.

In this current study we have investigated the use of colligation in the comparison of native learners of English and nonnative Pakistani learners of English. In this study, we have

explored the knowledge of colligation and extracted the different patterns of colligation in the comparison of native learners of English and nonnative Pakistani learners of English. To explore and investigate the use of colligation we have extracted the data from ICNALE (International Corpus Network of Asian Learners of English) The ICNALE has compiled from different native and nonnative learners.

In the ICNALE, all the three circles of Kachru model are covered. ICNALE has compiled from different native learners and nonnative learners. It is collected from learners data which are belonging to different parts of the world and ICNALE has provided the opportunity to get the data belonging all over the world from one place. ICNALE is the collection of the largest corpora which is comprised 2.3 millions of written essays. There are 2600 colleges in which we have collected data, and all these learners have taken from Asia. The ICNALE study has focused on the Asian learners.

ICNALE is the largest, valid and reliable source to get the data from the Asian learners (Ishikiwa, 2013) we have compiled two types of corpora. One is taken from the English native learners while on the other hand other is taken from Pakistani learners of English. We have compiled 400 files from Pakistani learners and 400 files from English native learners.

- All the learners have same level of education
- All the learners have same age group
- All the learners have same gender

The purpose of this current study is to investigate the syntactic behavior of non-native foreign language learners on the basis of corpus based learners data. (Luo & Zhou, 2017) and to promote the colligation by using different corpora. The initial analysis has proved that the native learners have used appropriate pattern of colligation whereas Pakistani foreign English language learners have not used the appropriate use of colligation. The frequent use of miss-colligation is mainly the result of unfamiliarity with colligation knowledge and the influence of the L-1 expressions.

The analysis has proved that the nonnative Pakistani learners have similar pattern like the native learners of English but they have different lexical behavior. Nonnative learners of foreign language have followed the same pattern of colligation but they have different lexical grammatical association like the native learners of English.

1.2 Role of Corpus Linguistics

Corpus linguistics is not the branch of linguistics it is the methodology. Methodology means to perform and corpus linguistics are used to perform investigating the problem, analyzing the problem and provided the solution of that problem.

To find out the linguistics features on the large scale for the quantitative research approach, the corpora has comprised on hundred, thousands, millions and billions words of texts rely on computer software which is used to determine the linguistic features. It is commonly used for both quantitative and qualitative analysis and the interpretation of qualitative results (Anthony, 2017).

In this research study we have the collection of large texts which are in the machine readable form and it has vital role in the field of the corpus. Corpus study has the quantitative measurement and quantitative measurement has scientific measurement (McEnery & Wilson, 2001, p. 29).

1.3 Difference in Frequencies

To explore the knowledge of colligation, we have drawn the comparison between the nonnative learners of English and native learners of English. In this comparison we have analyzed the both corpora which are balanced to each other. By comparing both corpora we have extracted the frequencies by using different corpus tools.

1.4 Pattern of Colligation

If we studied at the syntactic level, we have determined the different patterns at the syntactic level.

Sr No	Grammatical pattern	examples
1	Verb +to infinitive	(<i>he wanted to eat</i>)
2	object + bare infinitive	(<i>we saw it falling</i>)
3	Noun + wh-clause	(<i>The book which is on the table is mine</i>)
4	noun + possessive adjective	(<i>my brother is a good</i>)
5	Noun+ Modal verbs	(<i>Ali can do</i>)

1.5 Reason to use of colligation

Native learners of English have different frequencies to their learners and nonnative Pakistani learners have different due to their L1 background. The basic difference between Pakistani and native learners of English is the use of Grammar translation method which is the traditional method of classroom language learning.

1.6 Research Objectives

The aim of the study is to delve into the role of Corpus Driven Language Learning in English language teaching/learning and to make language teachers and learners aware and better able to recognize the advantage of corpus-driven language learning and to investigate the effectiveness of such teaching approach on the proficiency level learners.

The present study aimed to:

- Investigate the effectiveness of corpus linguistics in the study of colligation in between native English learners and Pakistani learners of English
- Examine the behaviour of Pakistani learners to learn English and their lexical grammatical attitudes towards the native learners of English
- Examine the Pakistani teachers and syllabus designers' attitude towards English competency

Research Questions

On the basis of quantitative research study(Cresswell, 2014), the present study endeavors to meet the objectives while answering the following close-ended ended questions

- What are the frequencies of the colligation in between the Pakistani learners of English and native learners of English?
- What is the difference between the patterns of colligations used by native and non native learners of English?
- What is the reason behind the use of erroneous colligation among non native Pakistani learners of the English?

1.7 Significance of the Study

The current study of the colligation has contributed by corpus based study and it is very important in the used of the colligation (Boulton, 2008).In this study researcher has explored the knowledge of the colligation and it is the occur of two lexical items which are grammatically committed to each other's.The nonnative learners which have not the proper use of colligation and they have not familiar with the pattern of colligation and the lexical behavior of the colligation are able to understand the proper use of the colligation different

pattern of the colligation are extracted and the strength of the colligation have investigate in this research study . In this research study ,teacher can familiar to the lexical behavior of the lexical items which are grammatically associated to each other's and teacher can taught the colligation which are minimized the difference between the native learners and nonnative Pakistani learners behavior of the colligation .

2. LITERATURE REVIEW

The semantic association between two words is called the collocation. Collocation is the set of the two words which occur together .Similarly if the set of the two words have the grammatical association to each other's that is called the colligation.. The words grouping to each other's and function on the syntactic level that is called the colligation (Nordquist, 2019) .The term colligation is not considered as the repeated words but also the classes of the words when occur at the same time and provided the habitual meaning in an utterance (Römer, 2005) .The word “colligation” is the Latin word and it has the meaning of the tie together. At the first time in the linguistic term this word has used by the Jhon Rupert Firth (who is the British linguistic). The Jhon Ripert Firth has defined the colligation as the interrelationship of the grammatical categories in the association of the different grammatical meanings at the syntactic level. The term colligation is the new way of the language learning which is based on syntactic level. The theory of the colligation has generated by the Michael Hoey work.In his book *Lexical Priming* he has described the way of languages learning and why we should be teach? Hoey has supported the idea of the Michael Lewis. This idea has been presented in the book of the *Lexical Approach* suggested by the Michael Lewis .This idea has suggested that it is the choice of the speakers that's he has selected the words and these words have determined the structure of the words that will use. This theory has argued that lexis is the complex and the systematical structure and the outcome of theses lexis is the grammar .this theory has reversed the rules of the grammar and the lexis.

2.1 Firthian Approach

As some distance as the Firthian technique is concerned, that means is taken into consideration as ‘a complex of contextual relation’ (Williams & Millon, 2012). Firth stresses at the vital position played by way of collocation in determining the means of a phrase, as may be seen from a well-known remark that “you shall know a word via the corporation it continues” (Firth, 1957, p.179). He explains the position of phrases in collocation is habitually positioned and the means of those phrases is jointly predicted and understood. Whereas, for colligation, he asserts that ‘that means at grammatical stage includes the interrelation of the kinds which are summary in nature’. Firth’s distinction between collocation and colligation is a useful one when you consider that each deals with which means at a syntagmatic measurement. However, which means of colligation operates at the grammatical stage, which isn't the same as the means of collocation on the collocational level. M.A.K Halliday and J.M.Sinclair in addition developed the Frithian technique. This Neo-Firthian technique to collocation places ahead a chain of concepts consisting of “lexical items”, “lexical set”, “node”, “span” and so on to set up a lexical principle independent of grammar. According to Halliday not all linguistics phenomena can be accounted for within the framework of grammar. The primary structural criterion is that of co-prevalence in any sequence, with or without intervening fabric: features which includes favored sequences or recurring intervention are secondary in shape(Sinclair, 1991). He, however, distinguished between ‘casual’ and extensive collocation based totally on some occurrences and manifestation of above hazard frequency respectively.

Collocations are pervasive in language. According to Altenberg (1991), “roughly 70% of the running words in the corpus form part of recurrent word combinations of some kind.” The investigation of such word combinations in corpora of authentic language dates back to the earliest studies of collocations by J. R. Firth (1957), who is commonly credited with

introducing the concept within British Contextualism. However, serious corpus-based exploration of collocations on a larger scale has only become feasible with the arrival of the computer in the linguist's workspace in the late 20th century. Since then, a substantial number of corpora of different sizes have become available, opening up new possibilities for collocation studies and many other linguistic applications. Progress has been made in particular by harnessing ever larger corpora, a growing range of statistical measures of association, and state-of-the-art software tools for automatic linguistic annotation and analysis.

“Semantically related lexical items” which are more commonly interpreted in the sense of semantic field relations such as “doctor – hospital – nurse” co-occurring within the same context. Their definition rests on a tendency of lexical items to occur in the same context because they belong to the same semantic field. The common underlying assumption is that collocations are characteristic co-occurrences of related lexical items, a notion that can also be identified in Eugenio Coseriu's (1967) concept of “lexical solidarity”.

Early approaches to the empirical study of collocation rest upon the manual identification of collocations in relatively small amounts of text (cf. Firth 1957), which are necessarily limited in scope and coverage. Since then, new research methods and tools as well as data have become available. The study of collocations has received fresh impetus through new computational approaches and the availability of large electronic text corpora, especially since the early 1990s. With the successively wider availability of ever larger corpora, studies of collocations have become feasible on a previously unknown scale, reaching a wider coverage of empirical data than ever before. Yet, in order to make such corpus-based studies possible and fruitful, the notion of collocation had to be not only defined, but also had to be operationalized. These new developments have brought about a further aspect in the definition of collocation, namely the definition and operationalization in terms of window-based approaches within the constrained context of a typically 3:3 or 5:5 key-word in context concordance window as proposed for example by Sinclair (1966, 1991). This approach has paved the way for the automatic investigation of collocations on the basis of relatively little linguistic pre-processing other than part of speech tagging, and by means of statistical methods modelling the characteristic co-occurrence of lexical items in terms of significance of co-occurrence as well as statistical measures of association.

2.2 Grammatical Approach of Colligation

Like Firthian and Neo-Firthian approach, grammatical approach also investigates the syntagmatic relations but inside a grammatical framework. As Lyons points out in 1966, sentences may be grammatically well formed, however they will be ‘simply nonsense’ in the event that they do not have the ‘implication of utterance’ from a semantic point of view (1966, pp. 291-292). There are 3 essential factors of semantic method:

1. The diploma to which lexical items are mixed with every other within the most obvious feature of collocations. Lexemes range noticeably with admire to the freedom with which they can integrate in syntagms with different lexemes. For instance, the adjective “appropriate”, “terrible” and many others. May be used as a modifier of any noun however “rancid” continually predicates the noun “butter”.
2. There exist restrictions among the lexical gadgets in collocation with the condition the lexical co-occur semantically. These regulations are semantically described in phrases of units including arguments and predicates (Wanner et al., 2013). For instance, “gold owns a man” is semantically anomalous sentence as “guy owns gold”.

“Encapsulation” is one characteristic of the connection between the lexical items in a collocation. If a lexical item is often co-happening with a restrained set of different lexical items, it could come to encapsulate their sense s(Lyons, 1977, p. 262). An example of such

encapsulation is offered via Lyons is the common collocation 2.7.1 Structural Approach to Colligation

This approach is based totally on grammatical structures, lexical items and the items with which they colligate. It has 3 most important points as discussed under:

Although supported Firth's concept, the more widespread Sinclairian use of colligation describes the co-occurrence of a category of grammatical items with a specified node. As an example, regarding the node true feelings, [John McH.] Sinclair notes that 'there may be a strong colligation with a possessive adjective...' other forms of colligation could be a preference for a specific verb tense, negative particles, modal verbs, participles, that- clauses, and so on. The notion that words may prefer (or, indeed, avoid) particular positions in text is picked up by [Michael] Hoey ([Lexical Priming,] 2005) in his more detailed definition of colligation: the essential idea of colligation is that even as a lexical item is also primed to go with another lexical item, so also it should be primed to occur in or with a selected grammatical function. Alternatively, it should be primed to avoid appearance in or co-occurrence with a specific grammatical function.(Hoey 2005:43) Hoey attributes his use of colligation also to check with sentential position as derivative from [M.A.K.] Halliday . . . ; it can, of course, even be seen as a natural extension of considering punctuation as a grammatical class, because punctuation is one among the foremost obvious indicators of positioning in text."(Gill Philip, *Colouring Meaning: Collocation and Connotation in Figurative Language*. John Benjamins, 2011.

"Colligation" is a term coined by Firth (1957) but little used since then. To some extent, its terms of reference have been superseded by those of the better-known term "collocation". If we take seriously Sinclair's assertion that there is no longer any sense in distinguishing between kids and grammar (1991:3), then the distinction between collocation and colligation to a large extent disappears. On the other hand, the term "colligation" is helpful in drawing attention to the fact that the evidence of many instances of naturally-occurring language can be used to explain behaviour that is traditionally associated with grammar. Just as the discipline called "kids" has been assisted by corpus-based approaches to collocation, so the discipline "grammar" benefits from corpus-based approaches to colligation. In his 1998 lecture, Hoey defined colligation thus:

- a) The grammatical company a word keeps (or avoids keeping) either within its own group or at a higher rank:
- b) The grammatical functions that the word's group prefers (or avoids);
- c) The place in a sequence that a word prefers (or avoids). This paper attempts to illustrate the role of colligation in clauses and in a text, relating this to the model of language description that Hoey expounded in his lecture.

The typical patterning of words, particularly in terms of the complementation patterns used with individual lexical items, has been studied by, among others, Sinclair (1991) and Francis (1993; 1995). The term "pattern grammar" has been coined to describe the lexical approach to grammar embodied in this kind of study (Francis, Hunston and Manning 1996; 1998; Hunston and Francis 1998; 1999). Distinctions in patterning are closely associated with meaning (Sinclair 1991; Lain 1995), to the extent that patterns can be used to distinguish.

2.3. Levels of Co-occurrence Restrictions in English Colligation

English colligation can be defined from unique degrees based totally at the co-prevalence restrictions. For the frame-work of present observe, sets of 4 degrees of collocational co-occurrence restrictions as proposed by means of Sinclair in 1996 [collocation, colligation, semantic preferences, and semantic prosody] and Allerton earlier in 1984 [syntactic restriction, semantic restriction, locutional restriction and pragmatic restriction], are accompanied. They provide a systematic way to explain the syntagmatic family members among lexical items, from concrete stage of actual collocation and grammatical policies to

the summary level of pragmatics. Both models are mixed collectively to begin the exam of actual collocates of a word its grammatical and semantic restrictions to its semantic prosody.

2.4. Sinclair Method

Sinclair proposes 4 degrees for the investigation of co-incidence of restrict: collocation, colligation, semantic preference and semantic prosody.

1. According to Sinclair, "Collocation is a frequent co-prevalence of phrases". Collocations are actual phrases and they can be immediately determined in textual statistics. Their probability of co-incidence can be statistically measured with the aid of t-score of MI-rating (Stubbs, 1995)
2. John Firth earlier has stated that Colligation refers to the interrelation of grammatical classes. Whereas, Sinclair refers to colligation as "co-incidence of grammatical alternatives". Hoey defines it otherwise, consistent with him colligation is "a grammatical agency word continues and the location it prefers". For example, there is a robust colligation for the phrase "accountant" with a classifier as in "a wages accountant" (Hoey, 2000).
3. "Semantic preference" is a lexical set of regularly happening collocates sharing decided on semantic capabilities (Sinclair, 1996; Stubbs, 1999). The lexical set is abstract and now not directly found in textual statistics (Stubbs, 1999). For example, a clear semantic preference for "true feeling" is "expression", that's inferred from the verbs with which "real feeling" often collocates "explicit", "talk", "show", "screen", "percentage" and so on. (Sinclair, 1996)
4. "Semantic prosody" suggests the audio system' mindset closer to what's being talked about and is at the pragmatic degree of language (Louw, 1993; Sinclair, 1996; Stbbs, 1999). This stage is likewise abstract and no longer observable directly. For example, Sinclair observes the placing characteristic of the phrasal verb "set in" within the nature of its subject this is constantly such as "decay", "ill-properly" or "prejudice" and many others. Typically refers to an unsightly situation. Stubbs in 1996 has in addition divided 'semantic prosody' into three classes: natural, high-quality and negative. As "set in" is an instance of terrible semantic prosody.

2.5. Allerton Method

Allerton (1984) has also advised four tiers of co-incidence of regulations: syntactic, semantic, locational and pragmatic.

1. "Syntactic regulations" refers to grammatical regulations between lexical objects in a collocation. If a speaker or creator fails to observe the grammatical rule which include an settlement among challenge and verb, a non-experience sentence could be produced.
2. "Semantic restrict" is the 'mutual semantic compatibility' of every phrase in a sentence, which constitutes its universal that means (Allerton, 1984). It is frequently the case for language inexperienced persons that they produce a grammatically best sentence but it sounds unnatural to native audio system. This phenomenon suggests that the that means of word-combos is not truly the sum of the meaning of every phrase: there should be some extra subtle sort of semantic phrases choice.

Semantic and grammatic restrictions are in addition distinguished under three members of the family: "tailoring", "reshaping" and "characterization" (Allerton, 1984). "Semantic tailoring" refers back to the fact that lexical collocates alter the which means of their companions. For example, 'a British Piano' would typically be assumed that piano is made in Britain; while, "a British pianist" want now not had been born there. The criterion judging Britishness is specific due to distinct collocates of "British" (Allerton, 1984). Semantic "reshaping" is the reinterpretation of a word due to its collocations. For instance, "fake enamel" are not real enamel but an alternative to them (Allerton, 1984). Semantic "characterization" refers to the phenomenon wherein an attribute calls for the presence of

unique characteristic in an entity it characterizes. For example, “the spinster’s husband” is atypical because of contradictory attributes of “spinster” and “husband” (Allerton, 1984). “Semantic characterization” is same, because the “contextual redundancy regulations” proposed via Leech (1974).

3. “Locution regulations” is referred to as “idiomatic” in Allerton’s model. It refers back to the linguistics phenomenon that the syntagmatic relations of lexical gadgets in this degree are arbitrary (Allerton, 1984, p. 28-36). For instance, the preposition in many verbal terms are arbitrary e-g “depression of”, “desire for”, “snicker at” and so on.
4. “Pragmatic restrictions” seek advice from ‘a lifestyle for who prefer one out of some of (near) synonymous terms or sentences (Allerton, 1984). For example, people favor to pick “decrease denomination expression” as ‘three hundred and sixty five days’ is desired to ‘a year’ (Allerton, 1984).

3.METHODOLOGY

This chapter explains the overall methodology undertaken in this research study.

3.1 Description of Data

This is the typical corpus based study in which we have taken two types of corpora and it is studied on the usages of the colligations in between Pakistani learners and native learners. In this study learners have focused on the lexical items which are tied together through grammatically.

The purpose of this study is the investigation of colligation pattern between Pakistani English learners and native learners of English.

It is basically corpus based comparative research.

3.2 Subject of the Study

Two types of data is required for the research study .ICNALE has provided the opportunity to attempt their analysis in between both Pakistani learners and the English native learners. 200 hundred files are extracted from Pakistani learners which are regarded the topic of smoking on the other hand 200 files are extracted from the Pakistani learners study part time jobs

Similarly 200 files are extracted from English native learners which are regarded as part time job and 200 files are extracted from English native learners and the topic is regarded as smoking.

All the files are extracted from the ICNALE wriiten essays

3.3 Comparison between both Corpora of Research Sample

- Both corpora are balanced with each other’s
- Both corpora belong to the same gender
- Both have the same level of undergraduate
- Both had taken same topics
- Both had the same age level

The information about their English learning background, learning strategies preferences, familiarity with different language learning techniques and skills were balanced and reliable data has considered in the research study.

Table 3.1 Data form ICNALE

ICNALE	400 files	400 files	Written essays
Pakistani learners	200 files	200 files	Written essays
English native learners	200 files	200 files	Written essays
Topics	Smoking	Part time job	Total 800 files

The data has collected from the ICNALE (International Corpus Network of Asian Learners of English) 400 files are collected from Pakistani learner's data and 400 files are collected from English native's learners.

3.4 Brief Outline

The ICNALE is an international learner corpus developed by Dr. Shin Ishikawa, Kobe University, Japan. The ICNALE includes more than 10,000 topic-controlled speeches and essays produced by college students (incl. grad students) in ten countries/regions in Asia (China, Hong Kong, Indonesia, Japan, Korea, Pakistan, the Philippines, Singapore, Taiwan and Thailand) as well as English native speakers.

3.5 Focus on Asian Learners

ICNALE has much focused on the Asian learners it has focused on both regions one is the ESL and other is the EFL regions. The European learner data has been collected in the corpora such as the international corpus of Learner English (ICLE)

The two types of data English for foreign language and English for second language learners have enabled to study on the different varieties of English not only the learners of the native speakers of English.

3.6 Collection of Varied Modes of Learner English

The ICNALE has consisted both mediums; spoken and written. Both mediums are important for the analysis. ICNALE has provided the option to the learners to pay the attention to difference between both monologue and dialogue.

3.6.1 Condition Control

The contrastive analysis of languages are truly reliable source of using the ICNALE. The purpose of ICNALE is the contrastive analysis of interlanguage (CIA). There are two common topics for both essays one genre is the spoken monologue and other genre is the written essays

There are two topics (a) it is important for college students to have a part time job and (b) smoking should be completely banned at all the restaurants in the country. Dialogue tasks are also related to those topics. The parameter of time for writing an essay, the parameter of length for writing an essay, the structure in the interviews, the time of the speeches are also controlled. ICNALE has provided the guarantees of using contrastive analysis.

3.7 Statistical Information of Data

The statistical information of ICNALE was obtained by wordlist tool of Antconc software. This tool provided us the total number of tokens and types from the both corpora as shown in table. There are 5266 types of the words in the English native learner's files while on the other hand there are 90613 word tokens in the e-files of English native learners.

To obtain this statistical information, we have used the Antconc and through word list tool we are able to know types and tokens.

3.7.1 Tools for Data Processing

In this section, we turn to the tools used to process the data for the present study: the PoS (parts of speech) tagger and suite of the Antconc tool.

3.7.2 Tree tagger tag set (Tag Ant)

Tag Ant is the software which has tagged data and this process is called tree tagger tag set.

3.7.3 Why Tagging is necessary

Tagging of corpora is necessary because it has the following advantages:

- syntactic patterns can easily be examined. As far as present study is concerned, six grammatical patterns such as "Verb + to infinitive" "object + bare infinitive" and "Noun + wh-clause" etc will be automatically obtained through the n-gram tool on Antconc software.

3.7.4 AntConc tools

In this research study, we have discussed the Laurance Anthony software tool of 64 bit.

3.7.5 n-gram/cluster tool

3.8. Framework of Methodology

Data was analyzed by comparison in between the native learners and nonnative Pakistani learners of English. The term colligation has been used in a large number of different senses. In fact, the use of colligation has been even more varied than that of collocation. The study framework attempts to synthesise recent advances in collocational theory into a coherent framework for lexicological theory and lexicographic practice. By posing a number of fundamental questions related to native learners and nonnative Pakistani learners of English and the definition of collocation, it critically reviews frequency-based, semantic and pragmatic approaches to collocation. It is found, among other things, that two types of collocation, namely 'long-distance' collocation and collocation between semantic features, have suffered almost total neglect. This leads to suggestions for a new division of the collocational spectrum and for a revised definition of 'collocation' based on the notions of 'usage norm' (Steyer 2000) and 'holisticity' (Siepmann 2003). It is argued that this new view of collocation considerably widens the dictionary maker's brief, since future lexicography will have to provide a full account of both structurally simple and structurally complex units, including fixed expressions of regular syntactic-semantic composition.

Ideas used for forming decisions and judgments

Adjectives Frequently Used With Framework

general or
 shared: **broad, common, comprehensive, general, overall, overarching, unified, unifying**
Within the broad framework of 'inclusive education', there are some clear underlying principles.
 clear: **clear, coherent, logical, robust, structured**
A clear and coherent policy framework is required.
 of a particular
 type: **analytical, conceptual, constitutional, contractual, ethical, fiscal, institutional, legal, legislative, methodological, national, regulatory, statutory, strategic, theoretical**
The legislation was aimed at harmonising the regulatory framework within which firms operate.

Verbs frequently used with framework

develop or provide a framework: **agree, build, come up with, construct, create, define, describe, design, develop, devise, establish, formulate, introduce, outline, present, propose, provide**
This constitution provided the first framework for democratic governance in the modern world.
 use a framework: **adopt, implement, use**
The paper attempts to implement a logical framework for the resolution of ethical issues.
 be a framework: **constitute, form, offer**
The charter contains six statements which form the framework of an overall policy. The 5-14 curriculum offers a framework within which such skills can be developed.

4. DATA ANALYSIS

4.1 Quantitative Data Analysis

The first type of data was analyzed quantitatively to reach the appropriate answer to the first two questions of the research study:

It is the quantitative study of colligation in which we are determined the lexical grammatical behaviour of colligation which is usually used by native learners of English and nonnative Pakistani learners of English.

4.1.1 Data Quantification

In the quantification of data, 400 native learners' files are extracted from ICNALE written essays and 400 files are extracted from Pakistani learners of English.

In the word list tool, we have determined the types and token in the native English learners files there are 5310 types of words and 276634 are words token which are used by native English learners.

According to the word list tool, in the Pakistani learners files there are 4838 types of words and token of words are 290875.

4.1.2 Pattern of Colligation

There are different pattern of colligation which are discussed in this research study colligation are not general term like a collocation .it is less general comparatively collocations study. In this study we have different pattern of colligation and their usages in the comparison of native learners and nonnative Pakistani learners.

<i>Sr No</i>	<i>Grammatical pattern</i>	<i>examples</i>
1	<i>Verb +to infinitive</i>	<i>(he wanted to eat)</i>
2	<i>object + bare infinitive</i>	<i>(we saw it falling)</i>
3	<i>Noun + wh-clause</i>	<i>(The book which is on the table is mine)</i>
4	<i>noun + possessive adjective</i>	<i>(my brother is a good)</i>
5	<i>Noun+ Modal verbs</i>	<i>(Ali can do)</i>

According to this above pattern of colligation, we are studied the comparison native learners and Pakistani learners of English.

In this study we are examined that the colligation patterns which are followed by nonnative Pakistani learners are closer to native English learners but frequencies of native English learners are usually low in the comparison of Pakistani learners because the nonnative learners are highly repeated the same colligation so Pakistani learners have high frequency of colligation.

4.1 Repitation of Colligation

Sr	Pakistani learners	Native learners
1	People who are	People who are

In this colligation pattern there 44 times repeated colligation in Pakistani learners of the study and 14 times repeated in the English.

4.2 Error nous of Colligation

Sr.	Pakistani learners	Native learners
1	Student is a big boy	This is not a big deal
2	Smoking is the big curse	College is a big enough
3	I am a big brother	I am a big fan

There are some lexical difference between the native English learners and Pakistani learners of English. In the use of colligation lexical items are used in different way. For example, at the above table the Pakistani learners have use the “big boy” in the sentence student is a big boy on the other hand native learners have used “big deal” they used the lexical item “big” with the word “deal”.

Native learners used the colligation like “I am big brother” while on the other hand nonnative Pakistani learners used the colligation like a “I am a big fan”. So in this study we have are determined to know that the native learners have used different grammatical lexicons while on the other hand nonnative learners Pakistani learners have different use of lexical items in the colligational behaviour of the study.

Verb + to infinitive

In the verb + to infinitive pattern we have investigated the some frequencies of verb which are comparatively higher in the native learners of the English but it is only done at the time when we have just single verb drawn in the search tool and investigate the frequency of verb it is comparatively high in the native learners but when we have extracted colligation in the between both learners corpus the colligational frequencies are higher in Pakistani learners due the repetition and have different lexical grammatical association between both Pakistani learners of English and native learners of English .

In this research investigation we have investigate the colligation in between the Native learners and the Pakistani learners of English.

Table 3.2 The pattern of colligations are below in the Table

We have drawn this comparison by using different research tools.

First of all we have tagged data via Tag Ant while tagging data we have catageroized the Pakistani data which are 400 files and files of native learners of English which are 400 files.

Sr No	Grammatical pattern	examples
1	Verb +to infinitive	(<i>he wanted to eat</i>)
2	object + bare infinitive	(<i>we saw it falling</i>)
3	Noun + wh-clause	(<i>The book which is on the table is mine</i>)
4	noun + possessive adjective	(<i>my brother is a good</i>)
5	Noun+ Modal verbs	(<i>Ali can do</i>)

Table 4.3Native and non-native frequency

Sr	Pakistani learners	Frequency	Native frequency
1	takes a large sum of money	1	1
2	takes a man to the grave	1	0
3	takes a part in managing	1	0
4	takes a person to painful death.	1	0
5	takes a person to painful death.	1	0
6	takes away their youth	1	0
7	takes from our side	1	0
8	takes only a few courses	1	0
9	takes out a smoke of cigarette	1	0
10	takes smoke in but the people	1	0
11	Takes a part in meaning	1	0

12	Takes a person to pain	1	0
13	Takes at least two hours	1	0
14	Takes away their youths	1	0
15	Takes at least two hours	1	0
16	Takes away their youth	1	0
17	Takes from our side	1	0
18	Takes out a smoke	1	0

Native learners have not used above colligations which are used by the Pakistani learners .In this table colligations are used by the Pakistani learners comparatively different way which is used by native learners of English .we have investigated the colligations which are started from concordance line of “takes” and investigate the colligation pattern of Pakistani learners of English which is different from the native learners of English.

In this balanced corpus we put the word “takes” in the search tool of Antconc and drawn the pattern of colligation which are located in the direction of takes .we have analysed the difference in the usage of colligation eg, Native learners have high frequency of the words “takes” but in the direction of the word “takes” they have low usage of colligation.

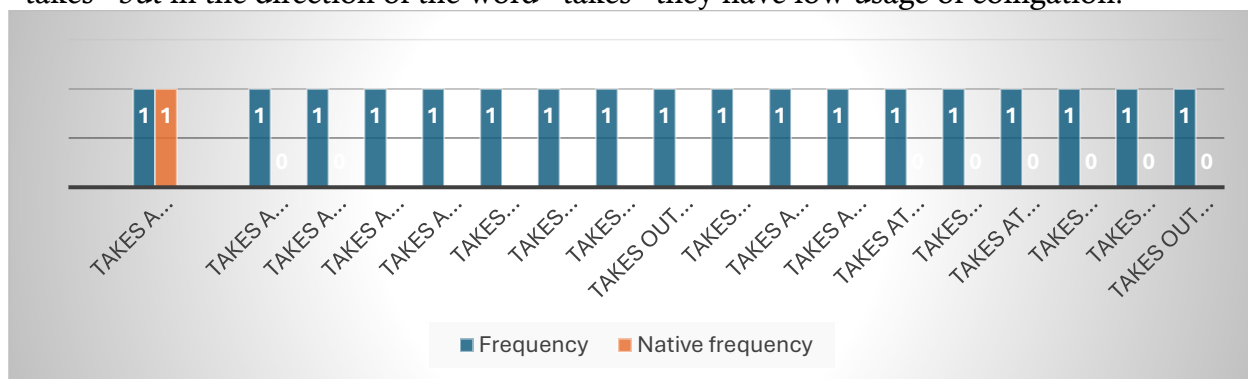
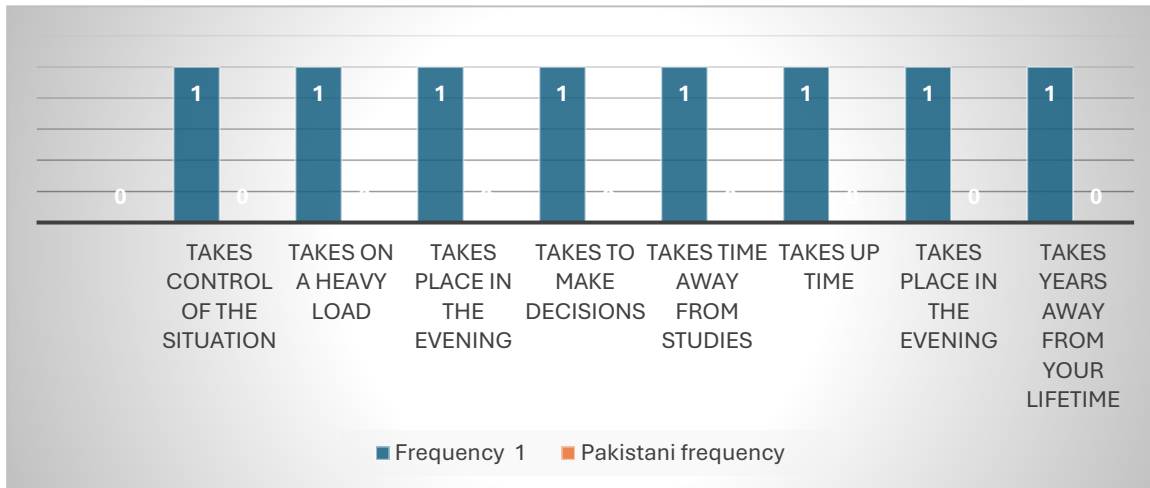


Table 4.4 Target colligation of native learners

Sr	Target colligation of native learners	Frequency	Pakistani frequency
1	takes to make decisions for themselves.	1	0
2	takes control of the situation	1	0
3	takes on a heavy load	1	0
4	takes place in the evening	1	0
5	takes to make decisions	1	0
6	takes time away from studies	1	0
7	takes up time	1	0
8	Takes place in the evening	1	0
9	takes years away from your lifetime	1	0



In this above table we put the word “takes” in the Antconc tool and this tool has shown the concordances of this word .we attempt it two times first for the native learners of the English and second time for the Pakistani learners of the English .

We examine the pattern of colligation which are associated with the word of “takes” .the frequencies of word “takes” in the data collected by Pakistani learners have 22 while on the others hand in the native learners have 24 .

After determined the words level frequencies we have investigatged the sets of colligation in the both of Pakistani and native learners.

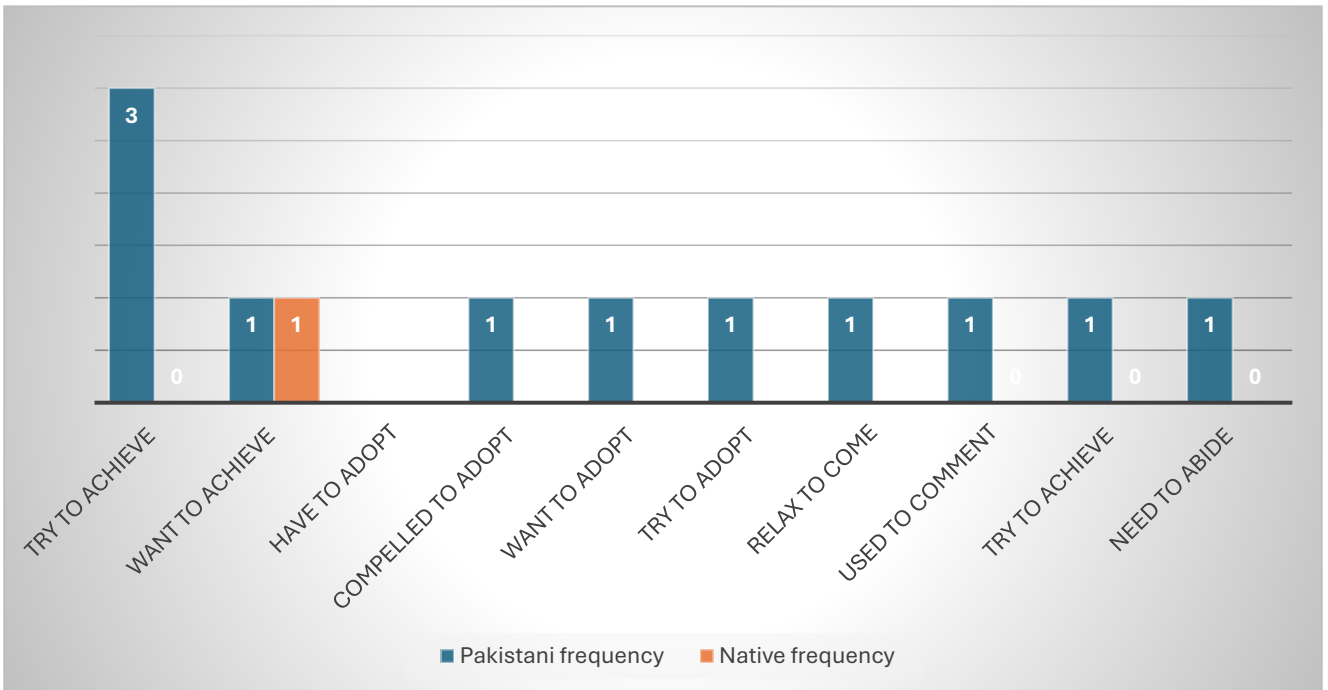
3.6.2 Pattern of Colligation

The above table of 1.1 has shown that the set of colligation in the Pakistani learners of English have higher than the native learners of the English.

Table 4.5

Verb +to infinitive

Sr	Pakistani learners	Pakistani frequency	Native frequency
1	Try to achieve	3	0
2	Want to achieve	1	1
3	Have to adopt	0	0
4	Compelled to adopt	1	0
5	Want to adopt	1	0
6	Try to adopt	1	0
7	Relax to come	1	0
8	Used to comment	1	0
9	Try to achieve	1	0
10	Need to abide	1	0



In the pattern of colligation, No doubt Pakistani learners have follow the pattern of in between native learners and Pakistani learners.

Table 4.10

Noun + Wh clause

Sr	Pakistani learners	Native learners
1	Problems who come	0
2	Families who come	0
3	Students who complete	0

Table 4.11

Sr.	Pakistani learners	Native learners
1	0	Smokers who come
2	0	Starbucks who ban
3	0	Husband who bring

In this pattern ,we have analyzed that the Pakistani learners have different colligation pattern in their study in comparatively of the native learners of the Pakistan.

Table 4.12

Sr	Pakistani learners	Native learners
1	Students who are	Students who are

In this colligation ,native learners have 9 times repeated this lexical pattern of colligation while on the other hand Pakistani learners have 19 repeated this lexical pattern of colligation

Table 4.13

Sr	Pakistani learners	Native learners

1	People who are	People who are
---	----------------	----------------

In this colligation pattern there 44 times repeated same grammatical lexical behavior of colligation in Pakistani learners of the study and 14 times repeated in the English native learners of the study.

At the above table 1.4 we have discussed some colligation pattern of Noun + wh-clause .In this pattern we have Antconc tool for both types of data and extract out the new colligation pattern of wh-clause.

The frequency list of English native learners have higher frequency level rather than Pakistani learners.

Native English learners have 477 frequency while on the hand Pakistani learners have 444 frequency of the colligation pattern.

Table 4.14

Sr	Pakistani Learners colligation	Pakistani frequency	Native frequency
1	Behind other actives	1	0
2	Do other activities	1	0
3	Interest in other aspects	1	0
4	Many other bad activities	1	0
5	Part other activities	1	0
6	The host of other disease	1	0
7	Ready to use other drugs	1	0
8	Compare to other jobs	1	0

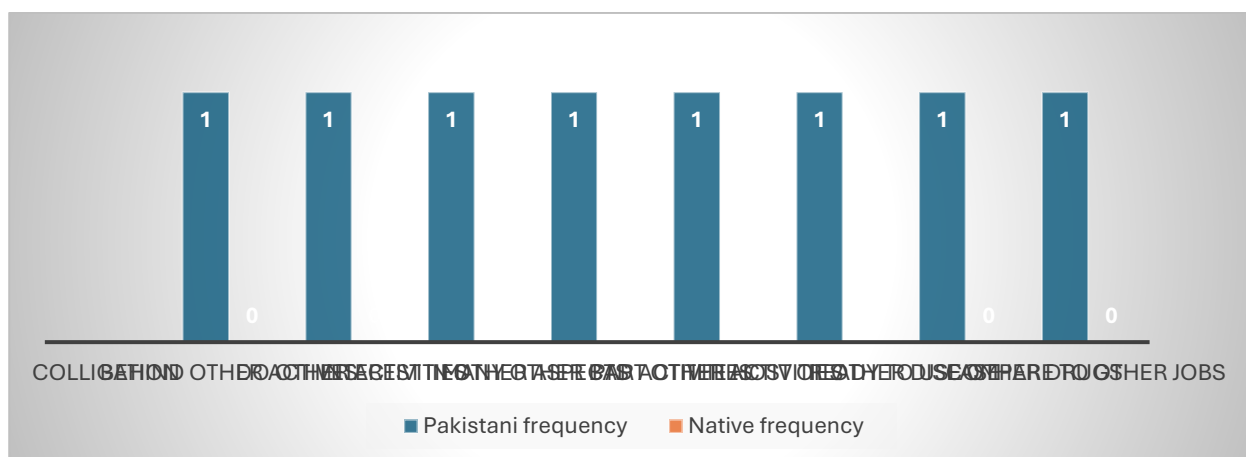


Table 4.15 In this table colligations are shown in table

Sr	Native learners	Sr	Sr
----	-----------------	----	----

1	Others to see	0	1
2	Others in danger	0	1
3	Others in determinant	0	1

Native learners have not used that colligations which are frequently used by Pakistani learners. Native learners have not used repetition in the use colligations.

In the above table of 1.5 we have put the word “other” in search tool of Antconc and the concordance list has shown .In this table we have extract out the less strength colligation and put the adjective “other”. In these concordance the Pakistani learners have higher frequency of adjectives while on the other hand Native learners have less frequency of less strength colligation.

In this table the frequency of native learners have 458 and Pakistani learners have the 766 .we drawn the less colligation pattern in between the Pakistani learners and native learners of English.

CONCLUSION

It has been concluded that the native learners of English have different lexical behavior of colligation in the comparison of Pakistani learners of English. In this research study, the problem has identified in the Pakistani learners and they have not used the knowledge of colligation appropriately. The data which is extracted from both learners have analyzed the frequencies and patterns of colligation in the Pakistani learners of English with the comparison of native learners. In this comparative analysis, we have studied different patterns of colligation and all these patterns have the different strength between the native learners of English and in the Pakistani learners of English. For example native learners have low frequency of repetition while on the other hand Pakistani non learners have high frequency of colligation. The same thing in all patterns of colligation are the equal syntactic pattern of colligation in between the native learners and nonnative Pakistani learners of English. But the major difference in between Pakistani learners and native learners are the lexical grammatical patterns which are native learners of English are used in the different way in the comparison of native English learners. We have analyzed the patterns of colligation and extracted that Pakistani learners have not used the balanced lexical items at the syntactic level. They have used different lexical behaviors of items. Moreover, Pakistani learners have the error to use the colligation and they have used inaccurate lexical items in the use of colligation. In the nonnative Pakistani learners the errornous of colligation has identified in the use of colligation and the reason is that the use of old traditional teaching methods and oftenly used the grammar translation method which has focused on the accuracy of the rules of grammar not on the lexical items. Due to this irregular behavior of lexical items, the errornous of colligation have identified in the writing of Pakistani learners. Pakistani learners have not perform the good strength of colligation in their writing essays and they have have not followed the accuracy in the use of colligation like the native learners due to their L1 background. Due to the influence of L1 background and attitude of syllabus designers, Pakistani learners have not used the appropriate lexical behavior in the used of colligation. So we can say in the study of colligation (in the comparison of native learners of English and Pakistani learners of English)has the similar pattern which is closer to native learners of the English but it has the different grammatical lexical association in the use of colligation between the native learners of English and nonnative Pakistani learners of English .

5.1 Pedagogical Implications

In the pedagogical implication, we suggest to improve the colligational lexical grammatical association in syllabus of Pakistani learners. Syllabus designers should to design the lexical

grammatical attraction like a native learners and teachers should taught the lexical item which are grammatical tied to each other's like a native learners at the initial grades

Native learners have the different lexical grammatical association but the nonnative learners have not used native like grammatical attraction at the lexical level so at the earlier stage only that syllabus has adopted which are designed according to native syllabus and teachers of Pakistani learners have taught that syllabus at the early grades so that learners can focused their lexical grammatical association like a native learners. The evidence in the corpora tends to indicate that learner English is produced in very distinct ways compared with those of the NSs. It is a characteristic of learner English to underuse these phraseological constructions or multiword units in the first instance. And when some details are under consideration, a tremendous amount of misuse in one way or another is found. In other words, the ICNALE written essays writers either do not produce the multiword sequences, or produce them in a deviant way. The formulaic nature of the English language is not appropriately represented in learner English. For many cases, ICNALE written essays writers are not found to be able to understand the colligational requirement and the semantic prosody of a particular pattern even though the forms are correctly produced as in the idiom of "might as well". In most of the cases, it seems that the learners have not produced a fairly amount of acceptable English when there is a strong demand for a particular colligation as in the pattern of "if NOUN must VERB ... then".

The fact that the sequence just as easily in the dictionaries within reach seems to point to the most immediate necessity to find out as many idioms as possible which may look loose but is commonly used by NS of English. The unveiling of this problem of the blend of two sequences with the same function is significant for pedagogy because it indicates a step missing after the two patterns: as ADV as NOUN can and as ADV as possible are introduced to students. It seems one further step is needed to bring the attention of the students to the structure of the two patterns and the necessity to disambiguate them. Only when this step is completed, can these two patterns be properly used without too much risk of mis-amalgamating them.

REFERENCES

- Altenberg, E. P. (1991). Assessing first language vulnerability to attrition. *First language attrition*, 189.
- Anderson, J. R. (1984). Cognitive psychology. *Artificial Intelligence*, 23(1), 1-11.
- Anthony, E. M. (2017). Learner autonomy in university English classrooms: Teachers' perceptions and practices. *International Journal of Applied Linguistics and English Literature*, 6(2), 19-25.
- Baleghizadeh, S., & Nik, A. (2011). The effect of type of context on EFL learners' recognition and production of colligations. *Dil ve Dilbilimi Çalışmaları Dergisi*, 7(1)
- Bartsch, S., & Holtz, M. *Corpus, Colligation, Register Variation*
- Berglund, Y. (2000). "You're gonna, you're not going to": a corpus-based study of colligation and collocation patterns of the (BE) going to construction in Present-day spoken British English.
- Biber, D., Douglas, B., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- Boldarine, A. C., & Rosa, R. G. (2018). Prepping a prep course: a corpus linguistics approach. *BELT-Brazilian English Language Teaching Journal*, 9(2), 379-394.
- Boulton, Z. (2008). The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL: the Journal of EUROCALL*, 26(2), 163.
- Butler, C. S. (2004). Corpus studies and functional linguistic theories. *Functions of language*, 11(2), 147-186.
- Cebik, L. B. (1969). Colligation and the Writing of History. *The monist*, 53(1), 40-57

- Celce-Murcia, M., & McIntosh, L. (1991). Teaching English as a second or foreign language.
- Channa, L. A. (2017). English in Pakistani public education: Past, present, and future. *Language Problems and Language Planning*, 41(1), 1-25.
- CHEN, J. S., & LIN, T. T. (2010). An Analysis on the Colligation and Collocation of the High-frequency Word Good in Learner Corpus. *Journal of Tianjin Foreign Studies University*, 1.
- Cheng, W. (2006). Describing the extended meanings of lexical cohesion in a corpus of SARS spoken discourse. *International journal of corpus linguistics*, 11(3), 325-344.
- Chujo, K., Oghigian, K., & Akasegawa, S. (2015). A corpus and grammatical browsing system for remedial EFL learners. *Multiple affordances of language corpora for data-driven learning*, 109-130.
- Coseriu, E. (1967). Lexikalische solidaritäten. *Poetica*, 1, 293-303.
- Cresswell, J. (2014). *Research Design (Qualitative, Quantitative and Mixed Methods Approaches)*. Book published by SAGE Publications 228.
- Dong, J., & Buckingham, L. (2018). The textual colligation of stance phraseology in cross-disciplinary academic discourse: The timing of authors' self-projection. *International Journal of Corpus Linguistics*, 23(4), 408-436.
- Ellis, R. (2015). *Understanding second language acquisition 2nd Edition-Oxford applied linguistics*. Oxford university press.