MASTERS THESIS

Study programme: Literacy Studies (MLIMAS)	Autumn semester, 2018 Open/~~Confidential~~
Author: Seyed Abbas Moazam Moosavi	………………………………… (signature author)
Supervisor: Professor Oliver M. Traxel	
Title of thesis: Investigating Frequency and Type of Lexical Collocations in Applied Linguistics Journal Articles Written in English by Iranian and Norwegian Scholars	
Keywords: Lexical Collocations; Applied Linguistics; Frequency-based Approach; Iranian Scholars; Norwegian Scholars	No. of pages: 94 + appendices/other: 4 Stavanger, 15 November 2018

ABSTRACT

In today's academic world, the research interest in corpus linguistics has shifted towards word co-occurrence rather than single words. Accordingly, a great body of literature has been devoted to investigations of recurrent word combinations in academic prose using frequency and dispersion parameters. This has resulted in analysis of corpus in different fields of study to collect comprehensive lists of academic collocations. Moreover, many contrastive studies have been conducted to compare the collocations used by native and non-native speakers of English. However, to the author's knowledge, few studies have been conducted to compare the most frequent collocations in two corpora of research articles written by non-native speakers of English published in international journals in the field of applied linguistics. To fill this gap in the literature, the current study investigated the most frequent collocations used by Iranian and Norwegian scholars in a corpus of 17 articles published in the *Journal of Pragmatics* through a frequency-based approach. Nine out of 17 articles were written by Iranian scholars including 67,673 words and eight out of 17 articles were written by Norwegian scholars comprising of 64,682 words. The data of this study were collected using *Collocation Extract* software. The results of the study were presented in three phases. In the first phase, 15 most frequent lexical collocations in both corpora were identified which were classified under three types of lexical collocations. Based on what was obtained, Adj+N collocation type had the most proportion in the corpora while Adv+Adj type had the least proportion. In the second phase, the lexical collocations of the Iranian corpus were presented including a total of 818 collocations classified under five types. According to the results, Adj+N was the most frequent type while N+V was the least frequent one. Similar to the Iranian corpus, lexical collocations of the Norwegian corpus were identified. They were classified under four types including a total of 462, among which Adj+N was the most frequent type while Adv+Adj was the least frequent one. In the third phase, frequencies of lexical collocations were compared in the two corpora. According to the obtained results, the two corpora did not have any had significant difference in the use of all types of collocation except for Adj+N type of lexical collocations.

Keywords: Lexical Collocations; Applied Linguistics; Frequency-based Approach; Iranian Scholars; Norwegian Scholars

ACKNOWLEDGEMENTS

Table of Contents

List of Abbreviations

EFL – English as a Foreign Language

ESL – English as a Second Language

ESP – English for Specific Purposes

L1 – First language

L2 – Second language

SPSS – Statistical Package for Social Sciences

List of Tables

List of Figures

1. Introduction

1.1. Overview

Since the time English became the lingua franca in international communication, many people all over the world have tried to learn English. Speaking English, as an essential skill, has played a vital role in communication between people from different nations. Besides speaking skills, which empower people to communicate orally, writing skills should also be learnt to facilitate communication through writing. The latter skill is not limited to daily routine communications but can be used in different discourses including academic writings, business letters and medical correspondences. However, it should be noted that each of these discourses has some specific features that may or may not be identical.

One of the discourses that has gained attention is academic writing. Authors should be able to use plain English in their writings in order for their paper to be accepted in professional journals. In doing so, researchers may face some problems in writing in a language other than their native language. Some of these problems may be due to grammatical issues, but most of them are usually related to features other than grammar. Corpus studies have shown that an immense part of both spoken and written language is composed of chunks, or various types of collocations and frequent word combinations (Sinclair 1991; Stubbs 2001). Evidently, these language features have gained the attention of different scholars in several fields of study (e. g. Sinclair 1991; Lewis 1993; Hsu 2007). The ability to combine words in the right way is of utmost importance to master any language and the key to native-like fluency. According to Lindquist (2009: 71), making minor mistakes in the choice of words by speakers of English as a second language with nearly perfect command of the language would mark them as non-natives. Therefore, the concept of collocations and appropriate usage of them is among the most interesting language features to the researchers.

Collocations are combinations of formulaic sequences of words. Collocations occasionally include non-compositional and opaque idioms, and in some cases, they are compositional and transparent combinations of words (Pawley & Syder 1983: 192). According to Sinclair (1991: 23), collocations are both stored and retrieved from the memory as single words and there is also no need for language users to try to formulate these

combinations based on syntactic rules. Additionally, these collocations play a vital role in the acquisition as well as rapid and fluent production of language. They help the producers to better express ideas and convey meanings. In fact, using formulaic language is not a matter of grammatical restrictions but it is a matter of convention. Normally, life-long exposure to conventions of a language makes their acquisition easy and conscious. That is why first language learners have no difficulty in acquiring these important aspects of language. But these conventionalized strings of words should be taught to foreign language learners due to their limited exposure to the target language. Moreover, the language-specific nature of collocations may compel learners to create combinations which are common in their first language but unconventional or impossible in the second language. Therefore, researchers who are writing in a language other than their native language may have little awareness about the use of these collocations in comparison to expert writers (Cortes 2004: 409). Since collocations are based on native language conventions, a lot of exposure to the target language is required in order to gain mastery of these linguistic features. Thus, the first language of the writers may affect their use of collocations in the target language. On the other hand, non-native writers who want to publish their articles in English-language journals should normally be aware of these collocations and their usage despite the effect of their first language. In addition, corpus-based analysis of recurrent word combinations indicates that variations exist in interdisciplinary fields and among non-native writers. Non-native writers may overuse, underuse or misuse English collocations or bundles in their writing or may have misconceptions of their pragmatic functions (Fan 2009: 110-123).

As pointed out, the ability to use correct collocations has been widely acknowledged in language acquisition and proficiency. Although there are many corpus-based studies that have investigated the use of multi-word combinations including three- or four-word bundles and their variations, usage of lexical collocations by expert writers has not been subject to much research. In this study, frequent collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars will be investigated. The study aims to fulfill the following purposes: 1. to investigate frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars; and 2. to determine the probable significant differences between the frequent lexical collocations used in research articles of applied linguistics written in English by Iranian and Norwegian scholars.

The thesis thus addresses the following research questions:

1. What are the most frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars?

2. Is there any significant difference between the frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars?

Research question no. 1 above is descriptive in nature; therefore, no hypothesis is put forward. For the second question, however, the following null hypotheses will be applied:

H0: There is no significant difference between the frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars.

1.2. Scope

As mentioned in the previous section, the present study attempts to find the most frequent English lexical collocations used by Iranian and Norwegian scholars in articles of applied linguistics published in international journals and to establish whether there is any observable difference between these two groups of scholars in their use of English collocations. Since there are a lot of articles in the applied linguistics field written by Iranian and Norwegian scholars that have been published in international journals worldwide, the scope of the choice of articles as the corpora of the study was limited by two factors. First, the articles were chosen from one academic database, namely *sciencedirect.com,* with the purpose of having consistent editorial policies and style preferences. Thus, the corpora include articles in sub-disciplines of applied linguistics that are written by Iranian and Norwegian scholars in English which are published in journals indexed in the ScienceDirect database. The second factor in choosing the articles was the date of publication. In order to eliminate the effects of time on the writing style of the scholars, it was tried to select articles which were published after 2000.

1.3. Background

Collocations are word combinations such as *extreme provocation*, *lay egg*, *cover entirely*, etc. "The term collocation has been labeled in a variety of ways, e.g. prefabs, multiword units,

etc. and defined in different manners in both linguistics and language teaching." (Fan 2009: 111). The only consensus is that collocation refers to "some kind of syntagmatic relation of words" (Nesselhauf 2005: 11). According to Benson et al. (1986)'s definition, fixed, identifiable non-idiomatic phrases and constructions are called "recurrent combinations", "fixed combinations", or "collocations which are between idioms" like *kick the bucket*, the meaning of which does not reflect the meaning of the parts, and free combinations in which one constituent can combine freely with other constituents. Benson et al. (1986: 7) divide collocations into two types: lexical and grammatical. Lexical collocations consist of the following combinations: verb + noun, verb + noun/pronoun, noun + verb, adjective + noun, noun + noun, adverb + adjective, and verb + adverb. Grammatical collocations consist of a verb, noun and adjective plus preposition or grammatical structure like a *that*-clause or an infinitive.

Most models of collocations converge in considering them as bipartite structures (Almela, Cantos, & Sanchez 2013: 231-240), "keyword/value" by Mel'čuk (1998: 23-53), and "node/collocate" in Sinclairian approach. Studies on collocations have been of interest to researchers after Firth (1957)'s lexical theory with its popularized slogan that "you shall judge a word by the company it keeps". He was the first scholar who drew attention to the fact that meaning is not restricted to single lexical units and disconfirmed the structuralists' view that words are presented in isolation. The same notion is emphasized by Stubbs (2002: 225), who conceives meaning as embodied in the "semantic relations between the node and collocates". In fact, polysemy of a word is disambiguated by its collocates and the arbitrariness of collocations posed problems to the use of synonyms in a collocational word pair (Mckeown & Radev 2000: 7). Thus, non-native production of infelicitous and unconventionalized language causes the speaker or writer to seem ridiculous and mocking despite the fact that non-native use of *powerful tea*, for example, instead of *strong tea* is meant by a native listener or reader as *strong tea*.

Moreover, the language-specific nature of collocations may compel learners to create combinations which are common in their first language but uncommon or impossible in the second language. For example, the collocations *dry bread* and *dry wine* in English are completely different from their counterparts in Persian. The problem will be further noticed when the structural difference plays a role in the sentence, for example in the sentence "They are constantly sending and receiving messages to and from the airport", particularly if the rendering happens from Persian to English.

Skehan (1998: 32) states that EFL learners rely on rules rather than lexicalized routines and they are pushed into combining words that do not normally go together. Most EFL writers with promising ideas are often not acquainted with the conventionalized collocations of the target language and create longer and wordier sentences, which results in unstandardized writing. Not only teaching vocabulary should be done through lexical phrases and not just isolated words, but also students should be exposed to the lexical bundles or collocations which are common in a specific field of study. By exposing EFL learners to any kind of probable word bundles, they will be able to commit these words to memory and define the semantic area of a word and as a result, as Nattinger (1988: 67) states, their predictions about the collocability of words will increase.

Being part of formulaic sequences (Wray 2000: 468), collocations are indispensable in ESL or EFL contexts (Brown 1974: 2). Brown (1974: 9) stresses the role of collocation in increasing EFL/ESL learners' oral fluency, reading speed, and listening comprehension. She was among the first few pioneers who proposed teaching collocations in classrooms. Nation (2001: 23) regards knowledge of collocations by EFL/ESL learners as a basic requisite to become fluent and achieve native-like proficiency in a foreign or second language. Furthermore, they perceived its role in producing appropriate language by learners. Hill (2000: 53) conceives collocations as an important aspect of vocabulary that help learners use words more frequently and proficiently. The dominant state of lexis has been reflected in Lewis' (1993) popular notion that "language consists of grammaticalized lexis, not lexicalized grammar", which emphasizes the interdependent relationship between grammar and collocations.

1.4. Relevance

Since this study was conducted on two corpora of articles written by scholars in applied linguistics, it conveys that both groups of writers, Iranian writers and their Norwegian counterparts, are graduates of or holders of degrees in linguistics, applied linguistics or other closely-related programs. Accordingly, this study can make a contribution to English language teaching programs and curriculums conducted in both Iran and Norway. At the theoretical level, identifying the most frequent collocations used in applied linguistics disciplines may result in a more comprehensive taxonomy of collocations which is particularly prepared for article writing in sub-disciplines of applied linguistics. At the

practical level, comparing and identifying differences in the use of collocations between the two corpora could help teachers recognize the weaknesses of the currently-used teaching curriculums. They can identify the problematic types of collocations and try to find or develop the appropriate methods of teaching those types of collocations in order to help students develop their skill in producing native-like collocations and be professional academic writers. The findings of this study will also be useful for those researchers whose aim is to publish their works in high-prestige international journals in the way that they become aware of the correct usage of collocations in order to write native-like articles.

1.5. Limitations and Delimitations of the Study

The limiting and delimiting factors of this study are source of corpora, time, wide range and extensive number of available articles which made the process of finding appropriate articles difficult. It was impossible to find a reliable source including classified articles in different fields of study based on the authors' nationality, mother tongue, and/or affiliation appropriate for the aims of this study. So, it was tried to choose the articles from a widespread academic database. Accordingly, in selecting the articles for the corpora, it was relied on the affiliation of the authors and the emails sent to confirm their mother tongue and their intention to authorize usage of their articles. Thus, another limiting factor was obtaining the authors' permission and consent. Individual emails were sent to the authors of the selected articles, but no reply was received from some of them. This also resulted in a waste of time waiting for their reply.

The other limiting factor that should be addressed concerns the quotations which are an integral part of all research articles in the academia. Although it was tried to omit almost all the quotations from the final corpora in the process of gathering and preparing texts, it was impossible to find and delete every single quotation particularly short and direct quotations in running sentences. This issue might affect the extracted list of collocations in the two corpora which results in slight deviations in the findings of the study.

Another limiting factor which could affect the results of the study was the date of the publications. Since writing styles could change over periods of time due to the changes in language, writing styles used in older articles could be fairly different from what is used nowadays. To overcome this obstacle, it was decided to select those articles which were

published after the year 2000. This eliminates the effects of time on the writing styles of the authors to some extent.

One of the delimiting factors of the study was related to the size of the corpora. Since a lot of articles were found which met the requirements of this study, the researcher had to delimit his choice of the articles. Therefore, it was decided to construct two corpora, one with articles written by Iranian scholars, and the other written by Norwegian scholars, each containing at least 60,000 words.

Another delimiting factor applied to the study was to select one reliable database in order to find the required articles and to have consistent editorial policies and style preferences. In doing so, *sciencedirect.com* was selected as the search engine, and *Journal of Pragmatics* was chosen as the source of articles to have consistent editorial policies and style preferences. However, delimiting the source made the process of finding desirable articles difficult since several different factors were involved in choosing an article.

Since collocations are divided into two main categories of lexical and grammatical collocations and each one is comprised of several different types, there would be a wide range of collocations to be searched for in the two corpora. Thus, the researcher decided to delimit the study to analyze lexical collocations which normally consist of two-word combinations.

The other delimiting factor was the comparison between two corpora. Although the aim of the study was to identify and compare the frequent lexical collocations between Iranian and Norwegian corpora, the researcher could derive no conclusion about the appropriateness and native-like usage of the collocations.

Moreover, since the topic of this study is narrowed down to identify the most frequent collocations in the Iranian and Norwegian corpora, the analysis of the obtained results was delimited to the main goals of the study. Consequently, the probable effect of L1 on the production of lexical collocations in L2 was not taken into consideration. Similarly, analyzing incorrect L2 collocations and probable errors in producing L2 collocations were not included in the scope of the present study.

1.6. Structure

This thesis consists of six chapters in total.

The current chapter provides introductory information in order to familiarize the readers with the subject matter and the aims of the present study. Then the researcher mentioned the research questions and related hypothesis. In the next sections, it was tried to shed light on the scope and background of the research. This was followed by familiarizing the readers with relevance and importance of conducting such a study and its limiting and delimiting factors.

Chapter two gives an account of the previous studies conducted in the field which are related to the theme and objectives of the present study. Accordingly, the theoretical studies – including trends, models and/or definitions – and empirical studies will be presented to shed light on the background of the study.

Chapter three presents the overall design of the study and introduces the research methodology including explanations about the research approach, the compiled corpora, research tools, the procedure of data collection and data analysis. This chapter ends with explanations of the ethical issues of the study.

Chapter four includes the presentation of the results of the study including the results of the statistical computations. Accordingly, the chapter starts with an overview of the research questions. Then, the results related to each of the two research questions are presented. First, the 15 most frequent collocations identified in both corpora are presented. Then, collocations found in the Iranian corpus are categorized based on their types. This is followed by the equivalent results related to the Norwegian corpus. Up to the end of this part the results are presented through descriptive statistics. Eventually, in the last part of the chapter, results related to the second research question are presented through inferential statistics.

In chapter five, 'Discussion' of the findings is presented. The first section of this chapter provides a summary of the findings of the study. Then, the discussion of the results is provided in the form of answers to the first and second research questions.

Chapter six is devoted to the 'Conclusion' of the study. In this chapter a brief conclusion of the results of this research is presented. This is followed by the theoretical and

practical implications, Then, limitations of the study and suggestions for further research are elaborated on.

2. Review of Literature

2.1. Introduction

In this chapter relevant literature to the subject of the study is presented. First, the subject of corpus linguistics will be introduced, followed by a description of different types of corpora. Then, various definitions of collocations will be given an account of and trends in collocations will be outlined. Finally, different empirical studies concerning collocations will be introduced.

2.2. Corpus Linguistics

Linguistics includes different branches such as sociolinguistics, psycholinguistics, etc.; these names by themselves are illustrative of what particular aspect of language they involve. But corpus linguistics is not illustrative by itself because "corpus" does not tell us what is under study. In fact, corpus linguistics is a methodology including several related methods which can be employed by scholars. This methodology is useful for those who are interested to study a language in use and to track its rules and changes (Lindquist 2009: 1).

According to Hidalgo, Quereda, and Santana (2007: ix), "corpus linguistics has played a powerful role in language research, grammar construction, dictionary making, natural language processing, cognitive studies, and language learning and teaching, among other fields". Corpus linguistics has also been useful in analyzing lexical patterns in a language. One of the aspects that corpus linguistics can be used for is to identify multi-word units used in a text to assess their frequency in a specific corpus. Hence, this helps the researchers to recognize and identify the most and least frequent sequences of the words in a specific text. This leads the researchers to make use of corpus linguistics in another aspect: language teaching. Researchers can examine how native-like students are and what their frequent errors are. Similarly, corpora can be used in textbooks to present authentic and real-life teaching materials (de Souza Hodne 2009: 28). Moreover, corpus linguistics can be employed to examine the effects of teaching on the production of specific parts of a language. It can be used to assess the effects of L1 on the production of the target language to compare the effects of two different L1 on the production of one target language. All of these goals

can be achieved by using appropriate corpora. The choice of the appropriate corpora shall be made in accordance with the purposes of each specific study. Using corpora has some limitations that one should be aware of. Hunston (2002) stated some of the restrictions of corpora:

> (1) A corpus will not give information about whether something is possible or not, only whether it is frequent or not. (2) A corpus can show nothing more than its own contents. […] Thus conclusions about language drawn from a corpus have to be treated as deductions, not as facts. (3) A corpus can offer evidence but cannot give information. […] The corpus simply offers the researcher plenty of examples; only intuition can interpret them. (4) Perhaps most seriously a corpus presents language out of its context. […] These factors all show the need for a corpus to be one tool among many in the study of language.
>
> (Hunston 2002: 22-23)

Therefore, choosing a proper corpus can help the researcher in achieving the designated goals of the study. Accordingly, knowing the definition of a corpus can give us a better view about corpus linguistics.

2.2.1. What Is a Corpus?

O'Keeffe, McCarthy, and Carter (2007: 1) defined corpus as a collection of texts which could be in written or spoken form that is stored on a computer. They believed that "a corpus can reveal the regular, patterned preferences of the language users represented in it, speaking and writing in the contexts in which the corpus was gathered" (2007: 60). Accordingly, they introduced some features of a corpus: "A corpus is a principled collection of texts", "A corpus is a collection of electronic texts usually stored on a computer", and "A corpus is available for qualitative and quantitative analysis" (O'Keeffe et al. 2007: 1-2).

By *a principled collection of texts,* they meant that any collection of texts cannot be attributed as a corpus. A corpus "must represent something and its merits will often be judged on how representative it is" (O'Keeffe et al. 2007: 1). They also believed that "a corpus is a collection of electronic texts usually stored on a computer" because this allows the researchers to access large amounts of stored text in order to be analyzed using various

specialized software. Regarding the analysis of texts, "a corpus is available for qualitative and quantitative analysis", which means that language features in a corpus can be analyzed in several different ways, i.e. qualitatively of qualitatively. In quantitative analysis, frequencies and digits are involved. On the other hand, in qualitative analysis, the notion of "how a word or phrase is used across a corpus" is analyzed (O'Keeffe et al. 2007: 2).

Not only deciding on the method of analysis of a corpus but also selecting the appropriate corpus strongly depends on the purpose of the research. And before choosing the proper corpus for any kind of research, one should know different types of corpora in order to achieve those specific research purposes. Thus, different types of corpora are presented in the following section.

2.2.2. Types of Corpora

As mentioned in the previous section, the type of corpora should be consistent with the aims they are used for. Not surprisingly, a large variety of corpora has been compiled by linguists and the number of corpora is growing every day. But it should be kept in mind that all of these growing corpora are usually categorized under the main types of corpora. Lindquist (2009: 11-22) described some of the main types of corpora which are briefly explained below.

Spoken Corpora

The main aim of the corpora is often to "represent general language at a particular point in time", though this type of corpora only covers certain types of language use under investigation (Lindquist 2009: 11). Spoken language is one of the interesting and important areas which is usually under-represented in general corpora because its compilation is complicated and expensive. An example of this type of corpora is The London-Lund Corpus of Spoken English (LLC) (Lindquist 2009: 11).

General Corpora

This type of corpora, unlike written-only or spoken-only corpora, contains both types of written and spoken data. Examples of this type of corpora are the British National Corpus (BNC), the bank of English (BoE), and Corpus of Contemporary American English (COCA) (Lindquist 2009: 15).

Specialized Corpora

The general corpora as described above can be used to investigate something about the language in general because they provide a picture of the language as a whole. However, in some cases general corpora are not practical for certain types of research questions and the researchers have to create specialized corpora to achieve their goals. Examples of this type of corpora are the Michigan Corpus of Academic Spoken English (MICASE) and the International Corpus of Learner English (ICLE) (Lindquist 2009: 18).

Historical (diachronic) Corpora

This type of corpora is useful for those researchers who are interested in changes of language over time and those who are interested in studying such a change over longer periods of time by comparing older texts with modern ones. Examples of this type of corpora are the Helsinki corpora and Lampeter Corpus of Early Modern English Tracts (Lindquist 2009: 19).

Parallel and Multilingual Corpora

The above-mentioned corpora contain texts in English; however, recently there have been a growing number of corpora which contain texts of two or more languages. This type of corpora is divided into two categories; some contain texts with their translation which are useful for research concerning translation, and some contain the same types of texts but in different languages. This type of corpora is normally used for comparative studies as well as translation (Lindquist 2009: 20).

Dictionaries as Corpora

"Electronic versions of dictionaries can be searched and used as corpora, but only if they contain authentic examples as illustrations" (Lindquist 2009: 20). Examples of such corpora are Dictionary of Old English (DOE), the Middle English Dictionary (MED) and the Oxford English Dictionary (OED) (Lindquist 2009: 22).

Text Archives as Corpora

Text databases are usually referred to as text archives but not corpora. This is due to the fact that these texts are collected for their own sake and are not balanced in any way. They can be used by researchers whose aim is to compile their own corpora. Examples of this type of corpora are Corpus of Late Modern English Texts (CLMET) and Corpus of English Novels (CEN) (Lindquist 2009: 21).

The Web as Corpus

This type of corpora, which has been developed recently, involves the compilation of data from the World Wide Web as a source for linguistic studies (Lindquist 2009: 22).

2.2.3. Selected Corpora for the Current Study

Since one of the aims of this study is to identify the most frequent lexical collocations used by Iranian and Norwegian scholars in a group of selected research articles, the type of corpora used in this study are specialized corpora compiled from widespread databases of research articles. The two corpora are specifically designed to meet the requirements of the study and include texts written by non-native speakers of English.

2.3. Collocations

There are many scholars who have defined collocations in various ways and it turned out to be an interesting topic to researchers and linguists. Most of these definitions are similar, but with slight differences. Firth (1957: 181) defined collocations of a given word as "statements of the habitual or customary places of that word". Firth looked at collocations as combinations of words and the meaning of these combinations was lexical meaning at the "syntagmatic level" (Firth 1957: 185).

Sinclair was another scholar whose definition was popularized. According to Sinclair (1991: 170), collocation is "the occurrence of two or more words within a short space of each other in a text". Sinclair (1991: 170) considered collocations as strings of words in which one word is called "node" and other words can come before or after this node to form collocations and these words are called collocates. Also, these combinations of words are used repeatedly so that they can be counted.

According to Manning and Schütze (1999: 141), "A collocation is an expression consisting of two or more words that correspond to some conventional way of saying things." They believe that collocations can be constructed by noun phrases, phrasal verbs, and other stock phrases. To them, collocations are identified by "limited compositionality" (Manning and Schütze 1999: 141). By "compositionality" they mean the meaning of an expression could be understood from the meaning of its parts. By using the word "limited" they mean

that collocations are not fully compositional because "there is usually an element of meaning added to the combination" (Manning and Schütze 1999: 141). So, in this manner collocations are different from idioms, which are non-compositional.

Later, Lewis (2000: 132) defined collocation as "the way in which words co-occur in natural text in statistically significant ways". He improved his definition in 2002 by adding that this phenomenon happens when "certain words co-occur in natural text with greater than random frequency" (Lewis 2002: 8). Following this definition, Nation (2001: 317) defined collocations as "items which frequently occur together and have some degree of semantic unpredictability". Also, O'Keeffe et al. (2007: 59) asserted that "Collocations are not absolute or deterministic, but are probabilistic events, resulting from repeated combinations used and encountered by the speakers of any language."

In addition, some scholars view collocations from a grammatical perspective. Kjellmer (1987: 133) defined collocation as "a sequence of words that occurs more than once in identical form in a corpus, and which is grammatically well structured". This definition implies that collocations are combinations of words which are defined at the lexical level and restricted at the grammatical level. By this definition it can be said that only a combination of two or more lexical words or a combination of one lexical word plus a function word sequencing identically is accepted as a collocation (Quping 2012: 31). From a grammatical point of view, Cowie (1994: 3169) also asserted that "collocations are associations of two or more lexemes (or roots) recognized in and defined by their occurrence in a specific range of grammatical constructions". Such a definition suggests that a random combination of words is not considered as a collocation unless it is grammatically well-structured.

On the other hand, Benson et al. (1997: ix) suggested another definition which considers collocation as both semantic and grammatical habitual combinations. This definition presented in the introduction of the BBI Combinatory Dictionary, is as follows:

> In English, as in other languages, there are many fixed, identifiable, non-idiomatic phrases and constructions. Such groups of words are called recurrent combinations, fixed combinations, or collocations.
>
> (Benson et al. 1997: ix)

From the aspect of vocabulary learning and teaching, Decarrico (2001: 6-7) described collocations briefly which are summarized in the following paragraphs:

Vocabulary knowledge is not only knowing single word or word families. It involves knowing the other words that normally co-occur with each other. These patterns, called collocations, are chains of two or more words that tend to occur with each other with high frequencies. This is also confirmed by Nattinger (1988: 69), who stated that "the meaning of a word has a great deal to do with the words with which it commonly associates". Therefore, knowing these particular pairs and groups of words plays an important role in vocabulary learning and word knowledge.

These associations and bundles of words help learners memorize the words and understand the semantic scope of a word. So, collocations should be learnt by L2 learners, if not, it results in some irregularities which mark the learners' writing or speech as non-native like, for example natives use *rancid butter* to refer to "spoiled" butter. This means that *rancid* suggests the collocate *butter*. But it should be noted that although *rancid* normally collocates with *butter*, the opposite direction is not as powerful. This means that *butter* weakly suggests *rancid* and *rancid* does not co-occur with other words but *butter* does. Therefore, the word that is restricted in the combination, e.g. *rancid*, is called the key word of the collocation. However, it should be noted that the key word is not always the first word of the combination.

One reason is that collocations are not always pairs of words such as Adj+N, they may occur as sequence of several words which could be three to five words long. The other reason is that the key word of a collocation can occur within a range of words and is not only limited to one word that it pairs. Therefore, a word such as *rancid* can occur with other words, e.g. *rancid lard, rancid oil*, etc. But it should be kept in mind that although the key words can co-occur within a range of other words, they are not allowed to collocate with any word, for example *rancid cheese* is not a collocation. Such restrictions may seem as an obstacle in the learning problems at first glance, but they normally help learners in their vocabulary skill.

One way of overcoming this learning obstacle is paying attention to the *semantic area* of the words. Regarding the previous example, *rancid* collocates with *butter, lard, oil*, etc. All these words have a common semantic feature; that is 'oily' as their base. Thus, teachers can use the semantic feature of the collocations to teach them effectively by presenting them in the context and raising the awareness of the student of the semantic links among them.

Concerning syntactic aspects of collocations, they are divided into two main syntactic groups: grammatical collocations and lexical collocations. In grammatical collocations, a word like a noun, a verb or an adjective collocates with a grammatical item, e.g. *by accident*. Unlike grammatical collocations, lexical collocations do not co-occur with a grammatical item, but they occur with other lexical items such as nouns, verbs, adjectives, or adverbs, e.g. *spend money*. Sometimes these types of collocations might be a source of errors because of the influence of L1. Therefore, teachers should identify the equivalent of problematic collocations and ask students to compare those collocations in L1 and L2 to identify the differences.

Fontenelle (1994: 44-5) presented a clear view of collocations from a slightly different perspective. He referred to Cowie's (1986) classification of collocations namely free (or open) collocations and restricted collocations. He explained that in free collocations one element of the collocation can be replaced by another word without imposing semantic change on the other element. For example, *eat* collocates with an infinite number of direct objects such as *cake*, *rice*, *chocolate*, etc. On the other hand, *eat* can also be replaced by other synonyms such as *devour, munch, gobble*, etc. Kuiper (2007: 97) believed that "restricted collocations involve preferential selection of word combinations where such combinations are arbitrary. They may also be idiomatic, i.e. not semantically compositional." Thus, in restricted collocations, "one element is used in a figurative or specialized sense", as the figurative use of the verb *blow* in "one can blow a fuse" (Fontenelle 1994: 44).

Fontenelle (1994: 44) again addressed Cowie's (1986) definition of overlapping collocations. As he explained that in overlapping collocations the first element, e.g. the verb *quench*, may collocate with two nouns, e.g. *fire* and *thirst*. While a verb, e.g. *distinguish*, can collocate with one of those two nouns, here with *fire* but not with *thirst*, another verb, e.g. *slake*, collocates with the other of those two nouns, here with *thirst* but not with *fire*. For restricted collocations, Fontenelle (1994: 45) referred to the examples of one sub-class of restricted collocations called delexical collocations. This type of collocation includes a grammaticalized verb and a direct object. The verbs in this combination "belongs to a closed class including highly frequent items such as *have, make, do, take, get, give*, etc.", for example "to give a sigh" or "to make a claim" (Fontenelle 1994: 45).

Fontenelle (1994: 45) believed that all the above-mentioned examples and definitions belonging to the free collocations are frequently classified as lexical collocations, as opposed

to grammatical collocations. He noted that grammatical collocations normally include "one element form an open class and an element from a closed class, typically, but not necessarily, a preposition." (Fontenelle 1994: 45), e.g. *depend on* is acceptable but not *depend of*.

Men (2017: 26) presented a useful summary of different definitions of collocation by various researches in the field. These definitions are presented in Table 2.1.

Table 2.1 Definitions of Collocations and Demarcating Criteria Adopted

Author	Definitions	Criteria
Aisenstadt (1979: 71)	"Combinations of two or more words used in one of their regular, non-idiomatic meanings, following certain structural patterns, and restricted in their commutability not only by grammatical and semantic valency (like the components of so-called free word-combinations), but also by usage"	Semantic transparency; commutability
Aisenstadt (1981: 54)	"A type of word combination consisting of two or more words, unidiomatic in meaning, following certain structural patterns, restricted in commutability not only by semantics, but also by usage, belonging to the sphere of collocations"	Semantic transparency; commutability
Van Roey (1990: 46)	"The linguistic phenomenon whereby a given vocabulary item prefers the company of another item rather than its 'synonyms' because of constraints which are not on the level of syntax or conceptual meaning but on that of usage"	Commutability

Howarth (1996: 47)	"Combinations in which one component is used in its literal meaning, while the other is used in a specialised sense. The specialised meaning of one element can be figurative, delexical or in some way technical and is an important determinant of limited collocability at the other. These combinations are, however, fully motivated"	Specialised sense of one element; commutability (collocability); semantic transparency (semantically motivated)
Nesselhauf (2005: 25)	"Combinations in which at least one element has a non-literal meaning (and at least one a literal one) and in which commutability is arbitrarily restricted, but some commutability is possible"	Specialised sense of one element; commutability
Laufer and Waldman (2011: 648)	"Habitually occurring lexical combinations that are characterised by restricted co-occurrence of elements and relative transparency in meaning"	Semantic transparency; commutability

Source: Men (2017: 26)

Considering all the above-mentioned definitions, it is evident that scholars do not agree on a common definition for collocations. Several factors must be taken into account to define this term. Among the various factors, lexical co-occurrence, grammatical and semantic factors have gained more attention. Thus, these trends in collocation are discussed in the following sections.

2.4. Trends in Collocation

Several studies (e.g., Kjellmer 1987: 133; Manning & Schütze 1999: 141; Lewis 2000: 132) have tried to explain English collocations. Most of these studies have focused on three trends, namely the lexical composition trend, the semantic trend, and the structural pattern trend.

2.4.1. Lexical Composition Trend

Advocates of the lexical composition trend assume that words gain their meanings from the words that co-occur with them (Gitsaki 1996: 136). Firth (1957) first introduced the term 'collocation' into lexical studies and is known as the father of this trend. He considers collocation as a 'mode of meaning' and a component separated from grammar. He states that there are four levels for the analysis of the lexical meaning: "the orthographic level, the phonological level, the grammatical level, and the collocational level." (Firth 1957: 192) Firth (1957: 195) had a "general rule" and believed that every word in a new context is considered as a new word. He also differentiated between collocational meaning and contextual meaning and divided collocations into two classifications of "general or usual collocations" and "more restricted technical or personal collocations" (1957: 195) without any further explanation.

Halliday (1966) and Sinclair (1966) were two advocates of Firth's trend who extended the theory and highlighted the role of lexical collocations as collocations that include lexical components. Sinclair (1966: 161) viewed language form by considering Grammar and Lexis as two "interpenetrating ways". Moreover, Halliday (1966: 148) believed that lexical theory is not a part of grammatical theory, but it is complementary to the grammatical theory. Sinclair (1966: 161) regarded grammar as a formula that forms language as a system of choices and the exceptions that could not conform to this system should be addressed to at the end of each grammatical description. On the other hand, for Sinclair (1966: 161) lexis was knowledge of individual lexical items and their tendencies to collocate which has nothing to do with grammar because they are more a matter of likeliness of occurrence than a matter of choice.

Advocates of this trend propose that this is the best way to analyze and examine collocation patterns which concentrate on the syntagmatic co-occurrence of lexical units (Alsakran 2011: 17). Gitsaki (1996: 141) stated that lexical composition trend had a good

point which showed that grammatical analysis was insufficient "to account for the 'patterns' a word enters in… and the collocatory idiosyncrasies of lexical items" so that it drew attention to lexis. According to the advocates of this trend, a lexical item cannot be described only by grammar; thus, it "must be identified within Lexis, on the basis of collocation" (Halliday, McIntosh & Strevens 1964: 35).

It should be reminded that advocates of lexical composition approach do not ignore the important role of grammatical analysis; they, however, tried to introduce an approach that was valid for lexical analysis without underestimating but considering the complementary role of grammar. As Halliday (1966: 159) admitted, scholars had no idea "how far collocational patterns are dependent on the structural relations into which the items enter", followers of this trend suggested that grammar was needed to help analyzing collocational patterns through lexical analysis.

2.4.2. Semantic Trend

The semantic trend tries to describe why words are combined with certain other words (Lehrer 1974: 178). The semanticists regard semantic properties of words as the basis for deciding which words can be combined with other words (Decarrico 2001: 7).

Historically speaking, Greek Stoic philosophers did not believe in one to one equivalence of meaning for each word, as Robins (1967: 21) asserted, they somehow considered the semantic structure of a language as an important aspect, so that "word meanings do not exist in isolation, and they may differ according to the collocation in which they are used".

Turning back to the current linguists, Chomsky was one of the first who treated collocations through a semantic view. According to Gitsaki (1996: 143), Chomsky differentiated "between 'strict subcategorisation rules', i.e. rules that 'analyze a symbol in terms of its categorical context', and 'selectional rules', i.e. rules which 'analyze a symbol in terms of syntactic features of the frames in which it appears'". Regarding Chomsky's categorization, selectional rules have a subsidiary role in grammar and they should be taken into account by semantics.

Lehrer (1974: 176) mentioned that lexical composition approach was not proper because it divided and categorized lexical items into categories based on their collocations, however it had no explanation for the question of why some words only collocate with certain other words. Lexical composition advocates analyzed collocations "as if the combinatorial processes of language were arbitrary" (Lehrer 1974: 176). Comparing to lexical composition trend, that analysis of language occurred at the lexis level and separated from grammar, in semantic approach, the collocations are analyzed based on a semantic framework, again separated from grammar (Gitsaki 1996: 142).

Due to insufficiency of lexical composition trend, Lyons (1977: 261), an advocate of the semantic trend, introduced the concept of "lexical fields" which is based on "the relations of sense holding between pairs of syntagmatically connected lexemes". In the strong version of field theory, vocabulary of a language should be consisted of a closed set of lexemes and each word should not belong to more than one field. Since vocabulary of a language is not a closed system in which lexis belong to various fields, this theory based on syntagmatic relations seemed problematic. Thus, Lyons (1977: 268) argued that descriptive semantics is sufficient without syntagmatic relations. Therefore, by ignoring syntagmatic relations, he started to deal with paradigmatic relations.

Moreover, there were other semanticists, who tried to propose a theory of lexical meaning with regard to the semantic features of lexical items, which means meaning of a lexis comes from the combination of the semantic features of that lexis. Here, the difference between lexical composition approach and semantic approach is that the former is based on a theory that considers lexis different from grammar but the later considers it complementary to grammar. Since in semantic approach, collocations are analyzed and defined by the semantic features according to its meaning or meanings, Lehrer (1974: 183) concluded that this approach may be more successful in explaining why some words just co-occur with some specific words but not with others.

As it was mentioned earlier semantic approach viewed the co-occurrence of words as a result of their semantic features. However, there are some criticisms regarding this trend because of the arbitrary nature of some collocations. According to Gitsaki (1996: 147), "there is a large number of idiosyncratic co-occurrences or combinations that are arbitrarily restricted", not based on their semantic properties. This was one of the weaknesses of the

semantic approach which resulted in problems in the way of analyzing collocations through a theory of lexical fields which were left unexplained by the semanticists.

2.4.3. Structural Pattern Trend

The structural trend includes studies that focus on the belief that collocation is affected by structure and hence collocational knowledge should be analyzed by considering its syntactic features (Hsu 2002: 42). Structural trend considers the importance of both lexis and grammar in the examination of collocations. It criticizes that "advocates of both the lexical and the semantic trends, examined a small set of lexical items due to their separation of grammar. Thus, their results were limited" (Alsakran 2011: 22).

Mitchell (1971: 48) pointed to the interdependence of words and grammar and asserted that 'lexical particularities' get their meaning from both contextual and lexical properties and the grammatical structure where they occurred. He suggested that collocations should be studied considering the grammatical structures (Mitchell 1971: 65). Greenbaum was also among the advocates of the structural trend. He admitted that the probability of co-occurrence of words and forming collocations should be dependent on syntax and believed that some certain words can only occur in certain structures (Greenbaum 1974: 82). Gitsaki (1996: 147) argued that without syntax, concept of collacability becomes meaningless, that is any two items can co-occur with each other but the acceptability of this kind of combination can only be evaluated through syntax.

Pawley and Syder (1983: 194) believed that language consists of chunks and blocks and one who wants to learn a language native-like, he/she should learn which of the grammatical sentences are native-like. Regarding this approach, learners learn a language in blocks, in fact they memorize the blocks. Nattinger and DeCarrico (1992: 92) also believed in this view, which resulted in compilation of a dictionary of English phrases for L2 learners. But, according to what they presented, lexical phrases are more general than collocations; moreover, they did not explain anymore about how lexical phrases are formed.

Kjellmer (1984: 162) presented some criteria to test whether a set of words occurring together is a collocation or not. He believed in studying collocations through the grammatical framework and defined collocations as "lexically determined and grammatically restricted sequences of words" (Kjellmer 1984: 163). Considering this definition, it can be said that

only those recurring sequences could be accepted as collocations that are grammatically well-formed.

Aisenstadt (1979: 71) also proposed the role of grammatical structures in the study of collocations and considered the restriction of the probability of collocations as a part of the extended notion of collocability. Combinations of words that their components are restricted within their 'commutability', are known as restricted collocations (Aisenstadt 1979: 71). According to this definition, restricted collocations are those combinations of two or more lexis employed in regular non-idiomatic meanings based on particular grammatical patterns that are restricted within their commutability not only by grammatical and semantic capacity, but also by their usage. He introduced the structural patterns of restricted collocations as V+(art)+(A)+N, V+Prep+(art)+(A)+N, A+N, V+Adv, and I(Intensifier)+A.

Later, three advocates of this trend, Benson, Benson and Ilson, compiled a dictionary for English collocations called BBI Combinatory Dictionary of English in 1986. It consists of more lexical items and does not attend to detailed grammatical and lexical explanations; it also does not include free combinations. The compilers defined and included 15 different sorts of "essential grammatical and lexical recurrent word combinations" for "general use" (Benson et al. 1986: 7). This definition divides the collocations into two comprehensive categories of lexical and grammatical categories.

For them, lexical collocations do not include grammatical structures, but they consist of nouns, adjectives, verbs, and adverbs (for more information and examples see Table 2.2 below).

Table 2.2 Types of Lexical Collocations

Types of Lexical Collocations	Examples
Verb + Noun	*to cancel an appointment*
Adjective + Noun	*strong tea*
Noun + Verb	*bombs explode*
Quantifier + Noun	*a swarm of bees*
Adverb + Adjective	*closely acquainted*
Verb + Adverb	*run rapidly*

Source: Adapted from Benson et al. (1997: xxx-xxxv)

On the other hand, grammatical collocations are those combinations that include a main word -such as verb, noun, adjective- and a preposition or a grammatical structure such as an infinitive or clause (for more information and examples see Table 2.3 below).

Table 2.3 Types of Grammatical Collocations

Types of Grammatical Collocations	Examples
Noun + Preposition	*Blockade against*
Noun + *to*-Infinitive	*a fool to do*
Noun + *that*-clause	*an agreement that s*
Preposition + Noun	*by accident*
Adjective + Preposition	*fond of children*
Adjective + *that*-clause	*afraid that*
19 different verb patterns in English e.g. verb + to-infinitive (they began to speak), verb + bare infinitive (we must work) and other.	

Source: Adapted from Benson et al. (1997: xxx-xxxv)

To summarize the three trends that were discussed in this section, it should be said that each of these approaches took different aspect of the phenomenon into account. The lexical composition trend considered lexical analysis as something independent from grammar in which lexis is an independent entity. "The semantic approach tries to find semantic features based on the meaning of lexical units that would enable the prediction of their collocates" (Gitsaki 1996: 161). Finally, the structural approach proposed a collocational pattern including both grammatical and lexical words.

Since this study deals with the comparison of lexical collocations used in two corpora, articles written in English by Norwegian scholars and articles written in English by Iranian scholars, the first view, lexical composition, is adopted. It should also be pointed out that six types of lexical collocations are searched for in the two corpora. These types are presented in Table 2.2 above.

2.5. Empirical Studies of Collocations in Applied Linguistics

Gitsaki (1996: 1-298) conducted a comparative study to examine the learners' knowledge of collocations in three tasks, namely essay writing, translation, and fill in the blank. For this purpose, ESL learners at three levels of post-beginner, intermediate, and post-intermediate were chosen. Findings of this study revealed a positive correlation between proficiency and the knowledge of collocations. Moreover, it was found that frequent collocation types were easier to learn by second language learners.

Moehkardi (2002: 53-62) discussed types of English collocations and the obstacles Indonesian learners could have in learning English collocations and some solutions to solve the problems of learning collocations. The researcher found that verb transitivity and phrasal verbs are one of the sources of difficulty in learning grammatical collocations. That is, they may confuse which verbs are transitive and which are not, which can be transitive under some situations, and "which structure (infinitive with or without *to*, gerund or *that*-clause) can follow certain transitive verbs" (2002: 58). Also, he addressed Verb+Adverb combinations as another source of confusion. Regarding lexical collocations, he mentioned that learners might feel more freedom in combining words to make lexical collocations due to the nature of lexical collocations. However, the problem he referred to in learning lexical collocations was the transfer of L1 elements and their influence on making lexical collocations. Finally, as a solution, he suggested that building up the learners' awareness to these types of word combination and helping the learners use them correctly and productively are some solutions that make learning collocations easier and more fruitful.

Hassanabadi (2003: 45-59) conducted a study to investigate the learning of lexical and grammatical collocations by Iranian EFL learners. He distributed a multiple-choice test consisting of 40 items among 80 EFL students of Shiraz University. The results showed that EFL students learn lexical collocations more easily than grammatical ones. They learnt the Verb+Noun category better than other categories of lexical collocations. He also found that Participle+Adjective+Preposition was the easiest to learn among the subcategories of grammatical collocations while Preposition+Noun was the most difficult one. The results also revealed that the degree of similarity or differences between L1 and L2 affects the learning of certain types of collocations. He finally concluded that special attention should be paid to teaching collocations, particularly the difficult ones due to the learners' weakness.

Nesselhauf (2003: 223-242) investigated advanced learners' use of Verb+Noun collocations. Thirty two essays of advanced German learners of English were collected through a free-writing task. Analyzing the data showed that learners' L1 had a great influence on their production of Verb+Noun collocations. It was found that wrong choice of the verbs were the most frequent collocational error type.

Jian, Chang, and Chang (2004: 4) described an algorithm which employs linguistic and statistical analyses to extract instances of Verb+Noun collocations from a very large corpus. Using the algorithm, they extracted valid instances instead of types, based on linguistic information of chunks and clauses. They also observed other types related to Verb+Noun such as Verb+Preposition+Noun and Verb+Noun+Preposition, which will be helpful in developing machine translation and computer assisted language learning.

Martyńska (2004: 2-12) also assessed the level of collocational competence among intermediate learners of the English language in Poland. The study was conducted among 53 high school students through a test on collocation. The results of the study indicated that learning individual words is not sufficient to achieve fluency in a second language. Therefore, in order to approach native-like level of proficiency, students must learn the words in chunks which are collocations.

Nesselhauf (2005: 1-275) intended to examine advanced learners' use of collocations. In fact, this study had for aims: 1. To identify difficulties of a group of advanced learners in the production of collocations; 2. To identify the factors that make these difficulties; 3. To identify the strategies and materials that learners use to create the collocations; and 4. To make suggestions for language teaching regarding the obtained results. The participants of were German-speaking learners of English. The researcher restricted the study to verb-noun combinations found in argumentative essays. The corpus of the study included 150.000 words of the learners' writings out of which a total of 2.000 verb+noun collocations were extracted manually. The results indicated that out of more than 2.000 verb+noun collocations found in the corpus, a quarter were wrong, and a third were deviant (which means they were wrong or questionable). Moreover, it was found that "the length of a learner's exposure to English in English-speaking countries was shown to probably have a slight effect on collocational accuracy, whereas the number of years a learner had undergone classroom teaching was shown to have no effect" (2005: 237). Thus, mere exposure improves the collocational performance to a slight degree. In addition, in the production of collocations, "neither

dictionary use nor time pressure seemed to have a significant effect on either the number of collocations produced or the number of deviations" (2005: 238). This indicates that learners were not aware of the problems that collocations made and they had no continuous control on their productions of collocations. Besides, it seemed that they lacked automatic control on the production of collocations and did not use collocations as much as native speakers to increase fluency if they wrote under time limit. According to the obtained results concerning problematic factors, "non-congruence between what the learner wishes to express in the L2 and the corresponding L1 expression was shown to lead to deviation in around 50% of the cases" (2005: 238). It was also revealed that there was a high number of word-for-word equivalence (over 60%) of V+N type of collocations in German and English. Another factor which resulted in collocation difficulty was the degree of restriction of a collocation. Additionally, the circumstances of production, the combination itself, and certain ways in which learners used collocation were correlated with deviations. It was found that L1 influence somehow occurred in half of the non-native collocations and dealt with all types of deviations. It was concluded that "characteristics of the individual learner, such as motivation, language aptitude, or preferred learning strategies, most likely also play a role" (2005: 246-247). Analyzing the way students produce collocations, it was revealed that 'transfer' was one of the strategies they used to produce collocations. However, it was mentioned that transfer could not be isolated from other processes in language production because what is produced is normally affected "not only by L1 but also by related L2 expressions' (2005: 253). Finally, it was found that "different types of transfer are of different strength for words of different word classes, or more specifically, verbs are much more often than nouns affected by transfer of formally related element" (2005: 252).

Nakata (2006: 154-168) compared two methods of learning collocations: meaning-focused and form-focused to investigate how meaning-focused and form-focused activities help learners develop their collocational knowledge. The results showed that there was a significant difference in the post-test scores of the groups taught by each of the methods. Form-focused methods were revealed to be more helpful and effective in the development of the collocational knowledge of the participants.

In a corpus-based study, Siyanova and Schmitt (2008: 429-458) detected production and processing of Adjective+Noun collocations in second language. Their corpus consisted of 1,810 Adjective+Noun collocations extracted from 31 essays written by Russian learners of English. Their analysis revealed that half of the employed collocations were frequently

used in the British National Corpus (BNC), whereas one quarter was not used in the BNC at all, and the other quarter had a very low frequency in the BNC. Accordingly, they concluded that although L2 learners were capable of producing a large amount of appropriate collocations, the fluency and underlying intuition of using collocations were not similar to native speakers even for advanced L2 learners.

Wang and Shaw (2008: 201-232) decided to test whether wrong collocations are due to transfer from L1. To this purpose, they compared the collocational errors of Swedish and Chinese students. A total of 100 university students participated in this research; they wrote a short essay based on the same prompt. Through analyzing the obtained data, the researchers identified the most common verbs and then occurrences of Verb+Noun collocations with the verbs *have, do, take* and *make*. They found that the two groups of the participants - Chinese and Swedish - had similar tastes in choosing sets of noun collocates; accordingly, they made similar types and proportions of errors. This implies that intralingual factors should be considered as important as L1 transfer in learning and using collocations.

De Souza Hodne (2009: 1-119) scrutinized teaching of collocations in upper secondary school level in Norway. She examined the vocabulary exercises included in textbooks of English taught at the first year of upper secondary schools in Norway. She tried to figure out that what portion of these words were part of collocations and among these collocations which ones were useful for Norwegian students in the first year of upper secondary school. She concluded that more than one third of the words included in the selected vocabulary exercises were part of collocations in the texts they were taken from and most of these words seemed to be useful for teaching.

Fan (2009: 110-123) employed a task-based approach in order to gain a deeper understanding of collocational use and the problems involved. The researcher investigated two highly comparable corpora: writings of Hong Kong ESL and native-speaking British students. The analysis of the results revealed that Hong Kong ESL students' use of collocations was not only affected by their L1, but it was also affected by their L2 as well as their inadequacy in the lexis and grammar of the target language. Therefore, the necessity for a pedagogical approach to the learning and teaching of this aspect of L2 is suggested by the researcher.

Chen and Baker (2010: 30-49) adopted an automated frequency-driven approach to investigate frequently-used word combinations in academic writing. They selected two

corpora of students' writings: one in first language and the other in second language. They found that a wide range of lexical bundles were used in published academic writing, whereas L2 students' writing exhibited a small range. The results also showed some overuse of some expressions in L2 students' writing while this was not the case with L1 students' writing.

Bhumadhana (2010: 1-130) tried to explore the most frequent types of collocational errors and the source of errors made by undergraduate students majoring in English at Walailak University. Besides, she compared differences in the use of verb collocations employed by three groups of low, moderate, and high English language ability students; and she examined the relationship between the use of academic verb collocations and writing competence of these students. She employed a writing test consisting of 21 items and distributed it among 155 students at three levels. It was revealed that the most frequent type of errors was Verb+Noun collocation, and approximation was found as the main source of error. The results also indicated that students with higher English proficiency gained a significantly higher average score on the test. And finally, a significant relationship was found between the writing ability of the students at the moderate level and their use of academic verb collocations.

Shokouhi and Mirsalari (2010: 1-24) investigated the probable correlation between the collocational knowledge and general linguistic knowledge of Iranian EFL learners on the one hand and identifying the hardest type of collocations to be learned by the learners. To achieve this goal, 35 participants in the study were examined using a proficiency test and a 90-item multiple-choice test consisting of both lexical and grammatical collocations. Analyzing the data revealed no significant correlation between the learners' linguistic knowledge and their collocational knowledge. Besides, they found that learning grammatical collocations was more difficult for the learners than the lexical ones. Moreover, Noun+Preposition type turned to be the most difficult and Noun+Verb type was the easiest to learn.

Alsakran (2011: 1-85) tried to examine the productive and receptive knowledge of lexical and grammatical collocations among advanced Arabic-speaking learners of English. The data were collected from a total of 68 participants: 38 Saudi students studying in Saudi Arabia, and 30 Arab students in the Intensive English program at Colorado State University. The methods used to measure the students' productive knowledge were three gap-filling tests: Verb+Noun and Adjective+Noun collocation tests as well as a Verb+Preposition collocation

test. Their receptive knowledge was also examined through a judgment test. The findings of this study showed that Arabic-speaking learners of English demonstrated poor knowledge of collocations on the four tests.

Concerning collocation studies in the Persian language, Darvishi (2011: 52-56) investigated the collocational errors in EFL college learners' writing. To achieve his goals, a total of 68 university students participated in the study. He gave the participants 38 assignments and 38 in-class practices to collect the appropriate data. He analyzed the collected data and identified unacceptable grammatical and lexical collocational errors based on the modified version originally proposed by Benson, et al. (1986: xxx-xxxv) and Chen (2002). The results showed that ignorance of rule restrictions was the major source of collocational errors. Also, interference of mother tongue, lack of the collocational concept, the interlingual or intralingual transfer, paraphrase and shortage of collocational knowledge were identified as other sources of errors.

Eftekhari and Rahimi (2011: 3941-3946) tried to investigate the effect of delexicalization of common verbs and level of proficiency on the collocational competence of Iranian EFL students. They conducted their study among 45 EFL students with low, intermediate, and high proficiency levels. They gave the participants a metalingual judgment test and asked them to judge the acceptability of 64 collocations of four common verbs (*have, give, take,* and *make*) in delexical uses in English. The results showed that knowledge of delexicalized collocations tended to fossilize at an intermediate level and did not increase with proficiency.

Concerning collocations in Malaysian English learners' writing, Hong, Rahim, Hua, and Salehuddin (2011: 31-44) employed a corpus-based error analysis to investigate the types and sources of verb-noun collocational errors in a subcorpus of a Malaysian learner corpus. The corpus included 130 essays written by Malay learners. Their findings showed that preposition-related collocations were the most frequent error occurring in the corpus. Moreover, intralingual transfer was identified as the most prominent source of collocational errors.

Laufer and Waldman (2011: 647-672) conducted a study to investigate the use of English Verb+Noun collocations in the essays written by native speakers of Hebrew. To this end, they compiled the corpus from the writings of the learners at three different proficiency levels. The corpus consisted of 3.000 words including argumentative and descriptive essays.

They also adopted LOCNESS as a corpus of young adult native speakers of English for making comparisons between native and non-native use of Verb+Noun collocations. Their results indicated that at all three proficiency levels learners used far fewer Verb+Noun collocations than native speakers of English. They found that the frequency of used collocations increased only at the advanced level. Analyzing errors revealed that errors, specially interlingual errors, continued to occur even in advanced levels of proficiency.

Bahardoust (2012: 185-200) sought to "evaluate the rate of lexical collocations in Iranian EFL learners' writing production across L1 and L2." She collected mid-term, final exam, and also the assignments of paragraph and essay writings of 200 Iranian EFL students whose native language is Persian studying at bachelor level. The findings revealed that the rates of Verb+Noun and Adjective+Noun collocations were the highest. On the contrary, the rate of Noun+Verb collocations was the lowest one. The results also showed that "L1 collocations were at higher frequency and rate" (Bahardoust 2012: 185) and L1 had both positive and negative effects on producing collocations.

Darabi (2012: 114-127) investigated "the possible relationship between receptive and productive knowledge of collocational patterns among Iranian TEFL university students." He conducted his research among 60 junior TEFL university students at Khorramabad Islamic Azad University, Iran. To achieve the goals of this study, a 70-item Collocational Behavior Test was used. The study revealed that the participants who were TEFL students had insufficient knowledge of collocational patterns. "It was also shown that there is a significant relationship between students' knowledge of receptive and productive collocation" (Darabi, 2012: 114).

Farrokh and Mahmoodzadeh (2012: 11834-11844) limited their study to explore Iranian learners' receptive and productive knowledge of English grammatical collocations of gerund in two proficiency levels. A total of 70 junior students majoring in English teaching participated in the study. They used three tests to collect the data. Their data analysis revealed that there was a significant correlation between students' receptive and productive knowledge of English grammatical collocations of gerund in both groups. They found that arbitrariness, unfamiliarity of subjects with English collocations due to insufficient exposure, and unfamiliarity with the rule concerning the use of collocations were sources of difficulty among the participants of the study.

Nejadansari and Alijanian (2012: 2-9) tried to explore the effects of pre-task and on-line planning on the production of collocations in speech. They employed a narrative task for 45 EFL students of a private language institute in Iran. The results of this study showed that on-line planning had a significant effect on the students' production of grammatical collocations though such an effect was not observed through using pre-task planning.

Shamsudin, Sadoughvanini, and Hanafi Zaid (2012: 1295-1302) also investigated Iranian EFL learners' collocational errors in speaking skill. The study was conducted on 15 Iranian postgraduate students studying in an Intensive English Course. The participants were asked to take two speech tests. Their data analysis revealed that the learners not only had a weak knowledge of collocations, but also their awareness of collocations had not been raised even to make them sensitive about collocations. Moreover, comparison of collocational errors in both tests showed that lexical errors occurred twice as many as grammatical ones. In addition, they found that the number of interlingual errors was much higher than intralingual errors. Finally, they concluded that teachers' emphasis on correct English collocations in the classroom can decrease the extent of collocational errors.

In a different study, Sharifi and Jafarpour (2012: 3-17) examined the effects of error correction feedback on various categories of lexical and grammatical collocations based on Benson et al.'s (1986: xxx-xxxv) collocations model. They conducted their study among 181 EFL students across elementary, intermediate and advanced levels of proficiency. Their study revealed that error correction feedback had a positive effect on the collocation competence at advanced and intermediate levels, while this effect was not observed at the elementary level. They also found that error correction feedback was better for grammatical collocation than lexical collocation patterns. Accordingly, they recommended that teachers use error correction feedback in teaching.

From a different perspective, Ahangari and Zununi Vahed (2013: 367-374) scrutinized types of collocations in Interchange series of books and high school books in order to investigate the differences between the two series of books regarding these types. They categorized the collected data based on the classification proposed by McCarthy (2005). The results revealed that pairs such as Verb+Noun were more frequent in Interchange book series while Adjective+Noun were the most frequent ones in high school books; though the findings showed insufficient use of collocations in high school books.

González and Ramos (2013: 563-570) tried to propose a method to evaluate the written production of Spanish collocations by Spanish learners. They compared two corpora, one written by native Spanish speakers and one written by Spanish learners, in terms of density, variety, sophistication, and number of errors. Their results showed that despite the use of collocations by learners, their choices lacked the variety, sophistication, and correctness exhibited by native speakers.

In line with the previous studies, Rassaei and Karbor (2013: 15-28) examined the role of three types of awareness raising techniques in learning a number of English collocations by Iranian EFL learners whose native language is Persian. These techniques were textual enhancement, input enrichment, and form comparison, which required learners to pay various levels of attention to collocations in the input. The findings showed that form comparison and textual enhancement techniques resulted in better acquisition of collocations but input enrichment technique did not.

Tekingül (2013: 1078-1089) investigated the effect of explicit collocation teaching on reading comprehension in comparison to explicit single-item vocabulary instruction in advanced EFL setting. She used a pre-test/post-test method to determine knowledge of third year English Language Teaching (ELT) department students before and after the treatment. The results showed no significant difference between the students of both groups, one taught by explicit single-item vocabulary instruction and the other by explicit collocation instruction. She concluded that one explanation for this result could be the prior knowledge of advanced level students; therefore, the teaching programs made no significant differences in the students' knowledge.

El-Dakhs (2015: 60-74) tried to assess collocational competence of Arab undergraduate EFL students, to figure out how collocational competence could develop by increased language exposure, to figure out whether word class of the collocates was a determining factor in the students' collocational competence, and to reveal the types of collocational errors they produced. She examined the collocational competence of 90 Arab EFL students through a specially designed test. She found that collocational competence of the students was not satisfactory; and this competence could be increased by exposure to the target language but with a slow rate. The results also indicated that the students used Verb+Noun collocations more confidently than Adjective+Noun collocations. Considering

the errors, it was revealed that intralingual errors occurred more frequently than interlingual ones.

Unlike most of the researchers interested in collocational studies, Park, Seraku, and Kiaer (2016: 1-25) had a different concern. Their purpose in the scope of collocations was not mere identification of the collocations of a corpus. Rather, they tried to establish criteria for collocations in Japanese and Korean to be distinguishable from free combinations and idioms. After scrutinizing the proposed empirical issues rather than focusing on statistical analyses they found that Im's (2006: 148–181) proposed criteria for Korean collocations could be employed for Japanese though with slight modifications.

2.6. Empirical Studies of Collocations in Fields Other Than Applied Linguistics

Gledhill (2000: 115-135) was interested in lexico-grammatical patterns in language and noted that these patterns such as lexical collocations, idioms, and phraseology of grammatical items could represent the prototypical phraseology of a particular genre. Thus, he tried to "describe the phraseology of the research article genre" (2000: 115) through a computer-based approach. He focused on collocations of grammatical words and took into consideration "the textual function of collocation and the role of fixed expression in the discourse community" (2000: 116). Gledhill (2000: 115-135) conducted his research using a corpus of 'introduction' sections of 150 cancer research articles. Analyzing the corpus, he concluded that collocations played different roles in the above-mentioned specialized texts. "In some instances, collocation involves terminology and reflects the recurrent semantics of the specialist domain. In other instances, collocation reveals the dominant discourse strategies in the research article" (2000: 130). He mentioned that these lexico-grammatical correspondences that were related to the cancer research article genre could be extended to different levels of specialization at different time periods. He finally concluded that "collocational patterns indicate a wider relationship beyond the individual text and reflect an evolutionary process that has forged the conventions of a number of phrases in the language of cancer research" (2000: 131).

Ward (2007: 18-35) tried to examine the relationship between collocation and lexical technicality and how this relationship can help teachers in teaching EAP to engineering students through a corpus-based study. Through analyzing the data, he tried to show that

formation of complex noun phrases (i.e. collocations) was a specific feature of engineering texts and such phrases were highly discipline-specific so that individual words did not have such potential. Then he argued that these highly discipline-specific collocations can be considered as a class of specific phrases in specialized engineering discourse and they can be used in EAP programs in terms of difficulty and specialization.

Ackermann and Chen (2013: 235-247) conducted a study to develop and evaluate Academic Collocation List (ACL). The corpus under examination included "written curricular component of the Pearson International Corpus of Academic English (PICAE) comprising over 25 million words" (Ackermann and Chen 2013: 235). The corpus consisted of written texts in different fields of study and academic disciplines. For applied sciences and professions, the corpus included texts from architecture, business, education, engineering, health sciences, media studies, and law. The texts related to humanities were derived from history, linguistics, literature, arts, general humanities, philosophy, and religion. In the field of social sciences, the disciplines were anthropology, archaeology, cultural studies, gender studies, politics, psychology, and sociology. For natural/formal sciences, the corpus included texts from earth sciences, chemistry, physics, computer sciences, mathematics, biology, and ecology. To develop the ACL, they followed four stages; first, they analyzed the corpus through computational analysis, second, they refined the obtained data from the previous stage based on qualitative and qualitative factors, then, the data were reviewed by experts, and finally, they were systematized. Since their aim was to develop a collocation list for academic purposes, they argued that although statistical information can help "identify and prioritize the corpus-derived collocational items" (Ackermann and Chen 2013: 235), it is only with the intervention of human that a data-driven collocation list can be of much pedagogical use. Using "a mixed-method approach of combining computational analysis of the source corpus with expert judgement and systematization" (Ackermann and Chen 2013: 246), they arrived at developing Academic Collocation List that can help EAP students and teachers focus on learning frequent collocations of their fields of study.

Gulec and Gulec (2015: 433-440) investigated the use of Verb+Noun lexical collocations across the health, physical and social sciences in the written academic genre. Their purpose was to find similarities and differences between the verbs with their collocations. They analyzed the texts using frequency and Chi-square tests. According to the results, they found 165 frequent verbs used across the three corpora. In addition, there were more similarities and relationship between health and physical sciences. However, social

sciences showed significant differences with health and physical sciences. It was revealed that collocates in social sciences had more variation in attaching to the identified verbs in comparison with collocates in the health and physical sciences. The number of verbs attracting collocates was more restricted in the health and physical sciences than in social sciences. Giacomini (2015: 140-148) conducted a corpus-based research on language for special purposes (LSP) collocational variation that she believed was mostly ignored in bilingual LSP lexicographic resources. She analyzed terminological variation on the formal and semantic levels. She chose a balanced monolingual LSP corpus including texts related to technical domain namely building and energy subfields. The corpus was collected through a specialized translation task at the Department of Translation of Heidelberg University. Analyzing this corpus, she tried to discover and depict recurrent patterns in collocational behavior. She finally found that present lexicographic resources mostly fall short to account for variational models in collocations.

León (2015: 526-534) was interested to identify the keywords and collocation strength in maritime texts in order to discover the terminological properties that make the maritime texts similar. To achieve this goal, she employed a corpus linguistics approach to analyze various types of texts of sister specialized sub-corpora in the maritime transport field. She used a contrastive analysis method to examine the frequency and coverage of register and to identify the distinctive features of the terms in each of the subcorpora although all were under the umbrella of semantic maritime field. Analyzing the obtained data, she concluded that variation and frequency of terms are closely linked with the communicative aims in various situations. She also added that "a term's keyness and collocation strength in a specialized language such as the ESP field of sea navigation is directly related to its level of technical semantic load." (2015: 534) She finally mentioned that the users' awareness of these collocations and keywords can help them to "confer language consistency to the different registers in the same domain." (2015: 534)

As the above-mentioned studies suggest, there are several studies on collocations that investigate various aspects of collocations from different viewpoints. As it was seen, some were concerned with proposing a collocational list in different fields or disciplines, some investigated the weakness and strength of EFL learners in using collocations to suggest methods for teaching collocations; and some examined the sources of errors and effects of L1 on the production of collocations in L2. However, no if any research has been devoted to identify and compare the most frequent lexical collocations used by Iranian and Norwegian

scholars in research articles written in English in the field of applied linguistics. Accordingly, the present study tried to identify the most frequent collocations in research articles written by Iranian and Norwegian scholars.

3. Methodology

3.1. Introduction

In this chapter, the procedure and methodology applied to conduct the present study are described and justified. According to the previous chapter, the aim of this study is to find the answer to the following research questions:

1. What are the most frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars?

2. Is there any significant difference between the frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars?

As the above-mentioned research questions suggest, the present study deals with the use of English lexical collocations in research articles written by Iranian and Norwegian scholars. Thus, the first step of this study was to find a reliable source of English lexical collocations, namely the *Oxford Collocations Dictionary*. Then, the corpora of the articles were selected. To this end, frequently-published international journals in sub-disciplines of applied linguistics were explored and chosen to collect the articles needed as corpora of the research. After finding the appropriate articles, they were converted into plain text format with the purpose of preparing the correct input for the software. In the next step, headings, journal names, author names, email addresses, affiliations, article dates, page numbers, footnotes, long quotations, tables, figures, references, non-English characters and other irrelevant text elements were excluded from the corpus. Afterwards, spelling errors due to conversion from PDF to plain text format were corrected.

Consequently, two distinct corpora were prepared, one for articles written by Iranian scholars and one for articles written by Norwegian scholars in internationally recognized journals. Then the two corpora were inserted into the *Collocation Extract* software (version 3.06) separately. First, the corpus related to Iranian scholars was added to the software and the procedure for running the software was conducted based on the instructions. Then the same procedure was conducted for the corpus of Norwegian scholars. In this way, the frequency of existing collocations was revealed through the output results of the software.

The obtained results were looked up and compared to what was included in the *Oxford Collocations Dictionary*. This enabled the researcher to identify the category of the collocations in order to exclude unacceptable collocations which are not in the lexical category. Finally, acceptable collocations were selected and listed for further data analysis. As the software did not observe capitalization of proper names in the texts, this issue was resolved manually.

3.2. Research Approach

In order to meet the goals of this study, a quantitative approach of research was applied. Using a quantitative approach allowed the researcher to obtain the detailed information required to identify the most frequent English collocations employed by both Iranian and Norwegian scholars and to compare the results in order to check whether they use the same or different types of collocations. But what is meant by "quantitative approach"? Since quantitative approach includes a wide range of theories, it should be mentioned that frequency-based theory was used in this research. Frequency-based theories of language analyze how frequency and repetition affect language which results in bringing about forms (Ellis 2008: 6). The frequency-based approach was first introduced by Firth (1957) and later developed by Halliday (1966: 148-162). Based on this approach, collocations are considered as "the combination of lexical items at a certain distance that differentiate between frequent and non-frequent collocations" (Alsakran 2011: 23). It should also be acknowledged that advocates of this approach are interested in "computational analysis of syntagmatic relations" (Nesselhauf 2005: 12). Since finding separated strings of words may result in low levels of validity and reliability, in this study non-separated collocations and collocations separated by one word are considered as data. Accordingly, this study takes a conservative approach in which a minimum frequency of 20 times per million words and an occurrence in at least 10% of texts are taken into account.

Most of the corpus-based studies investigating collocations follow this approach and the results are presented in descriptive format. Corpus-based studies are valuable because they provide the readers with an opportunity to identify the most frequent types of collocations or compare and find similarities or differences in the use of collocations employed by native and/or non-native English language writers. Thus, it enables the researchers in the field to propose implications for teaching advanced writing to English

learners for special purposes. Accordingly, the present study employs a corpus-based design to identify the most frequent collocations used in research articles written by Iranian and Norwegian scholars in international journals.

3.3. Corpora

A corpus is a body of text which is representative of a given language, dialect, or etc. used for linguistic analysis (Dash 2010: 1). By applying a corpus linguistics approach, researchers normally investigate language and its features through analyzing collections of text samples. But not every corpus is appropriate for all purposes. Selecting a particular corpus depends largely on the purpose of the research. Sometimes written language is the proper genre for a specific subject, other times a speech or spoken corpus. Likewise, the nature of the data is important in constructing a corpus. Based on the nature of the data, corpora are classified as general, special, sublanguage, sample, literary or monitor corpora.

Concerning the two corpora of the present study, they include articles written by Iranian and Norwegian scholars in English in sub-disciplines of applied linguistics. These articles are published in international journals and are chosen from the ScienceDirect academic database. Evidently, written corpora are used to find the answers to the proposed research questions. Moreover, these corpora are classified as specialized corpora because they are designed for a specific purpose and include texts written by non-native speakers of English. They have some features, including non-native authors and specific sub-disciplines, which make them inappropriate for other purposes. These two corpora are used for two purposes: firstly, to refer to the most frequent lexical collocations used by Iranian and Norwegian scholars, and secondly, to compare the two corpora in order to check whether there is any significant difference between these two groups of the scholars in their usage of English lexical collocations. Since a large number of articles meet the requirements of the corpora of the present study, it was decided to limit the time span of the publication date of the articles. Therefore, articles which were published between 2000 and 2017 were selected. Accordingly, the corpora of the present research consist of one corpus of articles in sub-disciplines of applied linguistics written in English by Iranian scholars and published between 2000 and 2017 and one corpus of articles in sub-disciplines of applied linguistics written in English by Norwegian scholars and published between 2000 and 2017.

3.4. Tools

Similar to other studies following a frequency-based approach, the present study uses some tools to achieve its goals.

3.4.1. Collocation Extract Software (version 3.06)

In order to achieve the goals of the present study, it was necessary to find the collocations used in the corpora. To do so and to decrease any human mistakes, *Collocation Extract* software (version 3.06) was used. This software is designed to identify and list collocations of a corpus; it can search the collocations of a word in the range of 2 to 5 words or all collocations of two-word chunks. Regarding the statistical methods, different inferential methods such as Dunning's Log Likelihood[1], Mutual Information[2], and Chi-square test[3] can be processed by the software. To run the software, one of the corpora must be added to the software and one of the statistical methods, for example 'raw frequency' if frequency of occurrences is concerned, must be selected. Then the span range from 2 to 5 must be specified. The number indicates the number of words to look for collocations. After that, it must be specified which options should be run. For example, if 'Left Side' is selected, the program looks for all collocates that occur before the keyword. Next, the minimum frequency of n-word collocations must be defined for the software. Also, the maximum items of collocations must be specified. Moreover, users can specify the distance between the two words. If set as '2', the two words are separated by one word. This option is provided because collocations sometimes can be separated by other words. Then the software could be run to search for the collocations.

---

[1] Dunning's Log Likelihood was first introduced by Dunning (1993). It is widely used as a measure of strength of association, especially lexical associations. It is better than Chi-square statistic X2 for dealing with rare events (Moore 2004: 1).

[2] "Mutual information (MI) is a measure of statistical dependence. The concept was introduced by Shannon (1948)" (Brillinger 2004: 163). Mutual information is one of many quantities that measures how much one random variables tells us about another. For more information visit: <http://www.scholarpedia.org/article/Mutual_information>.

[3] Chi-square test is one of the most frequently used measures to test the significance. "With the Chi-square test, you can test whether the measured difference in some respect between two groups is statistically significant or likely to be due to chance" (Lindquist 2009: 38).

3.4.2. Oxford Collocations Dictionary (2008)

According to the Oxford Learner's Dictionaries website[4], the *Oxford Collocations Dictionary* (2008) is a corpus-based dictionary that indicates which words work together and sound more natural. It helps students learn the most important collocations by choosing the correct combinations appropriate for specific contexts. This dictionary contains 250,000 word combinations, collocations for 9,000 nouns, verbs, and adjectives, and over 75,000 examples. Its corpus is based on the analysis of the Oxford English Corpus[5], a collection of nearly 2,5 billion words of English that show words which really do go together. The data obtained from the software were looked up in this dictionary in order to check the category of collocations and to confirm that the collocations are of lexical types. Thirteen types of collocations are introduced in the *Oxford Collocations Dictionary*, which are listed in the following table.

Table 3.1 Collocation types introduced in *Oxford Collocations Dictionary* (2008)

Entries	Types of combinations	Example
Noun entries	Adjective + Noun	*bright/harsh/intense/strong **light***
	Quantifier + Noun	*a beam/ray of **light***
	Verb + Noun	*cast /emit/give/shed **light***
	Noun + Verb	***light** gleams/glows/shines*
	Noun + Noun	*a **light** source*
	Preposition + Noun	*by the **light** of the moon*
	Noun + Preposition	*the **light** from the window*
Verb entries	Adverb + Verb	***choose** carefully*
	Verb + Verb	*be free to **choose***
	Verb + Preposition	***choose** between two things*
Adjective entries	Verb + Adjective	*make/keep/declare something **safe***
	Adverb + Adjective	*perfectly/not entirely/environmentally **safe***
	Adjective + Preposition	***safe** from attack*

Source: *Oxford Collocations Dictionary* (2008: ix)

---

[4] <https://www.oxfordlearnersdictionaries.com/definition/collocations>.

[5] The Oxford English Corpus mainly relies on the materials collected from World Wide Web. It includes all types of English such as literary novels, specialist journals, newspapers, magazines, blogs, emails, and social media from all over the world including the UK, the United States, Ireland, Australia, New Zealand, the Caribbean, Canada, India, Singapore, and South Africa. The corpus has nearly 2.5 billion words of 21st century English and new texts are being added continuously. For more information visit: <https://en.oxforddictionaries.com/explore/oxford-english-corpus>.

It should be mentioned that Quantifier+Noun and Verb+Verb combinations are excluded because the *Collocation Extract* software only shows bipartite structures or items. Grammatical collocations are also excluded as they are irrelevant to this study. Moreover, additional care should be taken into account in the analysis of Noun+Noun combinations depending on the role of the keyword of the collocation. Sometimes, this combination may be considered as a noun combination structurally while it is an adjectival combination semantically. This is due to the function of attributive nouns which are normally the keyword of the collocations.

An attributive noun is often placed before another noun and functions as an adjective. As Thomson and Martinet (1986: 13) explain, in cases such as *the walls of the town*, it is possible to omit *the* and *of*, and replace the first noun with the second one. The resulting combination would be *the town wall*s. In this case, the first noun becomes a kind of adjective which cannot be pluralized. Thus, the first noun that gains an adjectival function would be called an attributive noun. In fact, attributive nouns in English are singular, but as Bradley (1922: 112) states, it is better to say they are neutral, neither singular nor plural. According to Quirk, Greenbaum, Leech, and Svartvik (1985: 1333), attributive nouns are "normally number neutral […] nouns which are plural in post-modification are singular in pre-modification". That is in the first part of the combinations, a singular noun should be used even if the conception is plural. Therefore, to solve this problem and to ensure the type of identified combinations, the *Oxford Collocations Dictionary* was used.

3.5. Data Collection

Among the different methods of data sampling for a quantitative corpus-based study, namely random, regular, or selective method, random sampling seemed to be the most proper method for the present study because it is a reliable technique which is widely applied in many disciplines of natural and social sciences. Therefore, in order to find and compile the corpora of the present study, the random method of data collection was utilized. In doing so, one of the most widespread academic databases, i.e., ScienceDirect, was considered as the main source for finding the target corpora. In order to save time and energy, a few keywords such as 'applied linguistics', 'Iranian', 'Persian', 'Norwegian', etc. were used. The results suggested by the search engine of the above-mentioned database were checked to see whether the articles are appropriate. In selecting the appropriate articles some factors were considered.

First, the articles should be written in applied linguistics sub-disciplines. Second, the name and affiliation of the scholars should indicate that they were either Iranian or Norwegian. Third, the articles published after 2000 were selected to be included in the corpora. Then, to ensure the nationality of the scholars, emails were sent to the scholars and only those articles whose authors replied to the emails were included in the corpora. In order to have a better understanding of the process of preparation of the corpora and data collection of the current study, a sample of the research article author's confirmation email is presented in Appendix 1.

Consequently, a total of nine articles including 67,673 words written by Iranian scholars were selected. Similarly, a total of eight articles including 64,682 words written by Norwegian scholars were selected. These articles were published in the Journal of Pragmatics and on the ScienceDirect website. Lists of the articles used in both corpora are presented in Appendix 2.

3.6. Data Analysis

Linguistics as a general subject is closely related to statistics and mathematics. Hence, statistical and quantitative methods play an important role in conducting research in any sub-fields of linguistics such as corpus linguistics, computational linguistics, applied linguistics, etc. Thus, knowing different properties of a language and having a good knowledge of statistical information about these properties can help one in using linguistic data (Yule 1964: 10).

As it was mentioned in section 3.2, the present research follows a quantitative approach. Therefore, it applies quantitative analysis to analyze the data. Generally, in quantitative analysis different linguistic properties of a language are classified and counted to construct statistical models in order to explain what is observed by the data. This enables the researchers to find out which phenomena are manifestations of the features of a language or they have just occurred by chance (Dash 2010: 12).

There are different statistical approaches to conduct quantitative analysis of the data such as descriptive statistical approach, inferential statistical approach, evaluative statistical approach, multivariate statistical techniques, etc. By descriptive statistical approach, one can summarize the most significant features of the examined data. Inferential statistical approach

enables the researchers to answer the questions with much certainty through using descriptive statistical methods. Evaluative statistical approach is used to "test whether a hypothesis is supported by evidence in data, and how the mathematical model or theoretical distribution of data relates to reality" (Oakes 1998: 1). Also, multivariate statistical techniques are used when it is planned to make comparisons and hidden patterns from raw frequency data (Dash 2010: 12). Thus, to achieve the objectives of a study, any of these methods can be used according to the nature of the research goals and questions.

Regarding the research questions of the present study, which are mainly descriptive in nature, it is decided to use 'descriptive statistical approach' to answer the first question and 'inferential statistical approach' to answer the second question. First, the two corpora were added to the software. After running the software, an output of a raw list of collocations was obtained. All the data obtained from the raw list of collocations were compared with the *Oxford Collocations Dictionary*. This was done for both corpora. Then, the results of the comparisons were put into statistical analysis. They were entered into *Statistical Package for the Social Science (SPSS)* software and analyzed using descriptive statistics in terms of percentage. Further, to detect any probable differences in the use of collocations between the two mentioned corpora, inferential statistics in terms of Chi-square tests were employed.

3.7. Reliability

In this section the aspects related to the reliability of the current study are presented. The first issue that should be addressed is the process of selecting the proper articles for the corpora. Since there was no specific corpus including articles written by either Iranian or Norwegian scholars in sub-disciplines of applied linguistics, the researcher had to find the proper articles. To achieve this aim, it was tried to use only one of the available academic databases, which was the ScienceDirect website. This limits the range of journals and increases the probability of analogous articles by both groups of the scholars presented in the same journals, which itself increases the reliability of the work.

Another issue which should be mentioned is that not every co-occurrence of the words was considered as a collocation. To avoid software mistakes, the output results were looked up in the *Oxford Collocations Dictionary* and were compared with what was

presented in this dictionary. Those outputs which did not match the collocations of the *Oxford Collocations Dictionary* were excluded from the results.

Besides, the process of comparing the output results of the software with entries of the *Oxford Collocations Dictionary* was conducted by the researcher and an assistant. By doing so, it was tried to decrease any human mistakes. If there were any differences in the results corrected by the researcher and the assistant, those items were discussed to resolve the problem.

3.8. Ethical Issues

By the advent of online publications, ethical issues have become complicated. Before the age of online publications, it was necessary to get the printed copy of the articles from the publishers and make the corpus; but nowadays this has become much easier simply by downloading the articles from the internet. However, the researchers encounter two important ethical issues here. First, whether they should have the scholars' permission, and second, whether they can reproduce and distribute the corpora or in other terms how to deal with the laws of Copyright.

To address the first issue, the researcher decided to get permission from the authors of the articles. So, he sent emails to the scholars and described his work and asked them about their nationality and their tendency to participate in this particular research. Then, those articles whose authors replied to the emails and were eager to cooperate were included in the corpora. One of the confirmation emails sent to the research article authors, is presented in Appendix 1 as a sample.

In order to solve the second issue, the researcher decided not to publish the whole corpora in plain text format. Instead, he decided to refer to the articles by preparing a list of articles used in preparing each corpus which is presented in Appendix 2 of the current study.

3.9. Summary

The present study is a quantitative type of research which follows two goals: 1. To find the most frequent types of English lexical collocations used by Iranian and Norwegian scholars in articles in sub-disciplines of applied linguistics; and 2. To find out whether there is any

significant difference between the frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars. Like any other piece of research, this study has some particular features which were addressed in detail in the previous sections of this chapter. All the reasons for choosing a particular approach, for preferring one method over another, etc. were provided in this chapter. Two tools were also used to analyze the data, *Collocation Extract* software (version 3.06) and *Oxford Collocations Dictionary* (2008). Moreover, issues related to the reliability and ethics of the present study were explained in relevant sections. It should also be mentioned that every step taken by the researcher to conduct this study is described and clarified so that the procedure of the research could be easily traced by the readers.

4. Results

4.1. Overview

This chapter tries to find the answer to the following research questions through statistical methods and analysis. These questions are as follows:

1. What are the most frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars?

2. Is there any significant difference between the frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars?

To address the above-mentioned questions, this chapter presents the results obtained from the analysis of the data collected from the two corpora. For this purpose, two statistical methods are applied, namely descriptive and inferential statistics. Accordingly, the results are provided in four phases; the first three phases are related to the first question and the fourth one deals with answering the second question. The first phase presents the descriptive analysis of the data obtained from the two corpora in terms of frequencies ─ illustrating the 15 most frequent collocations used in Iranian and Norwegian research articles. In the second and third phases, the descriptive statistics of the results of the used collocations in Iranian and Norwegian research articles are presented. The results are shown in terms of frequencies and percentages. Finally, the fourth phase is devoted to the comparison of common collocations used in both corpora through inferential statistics, namely Chi-square tests.

The procedure used to identify the frequent collocations was based on the minimum frequency of 5 times of occurrence in the corpora because it is the minimum frequency which can be calculated and is valid using statistical tests such as Chi-square and Dunning's Log Likelihood test.

4.2. Fifteen Most Frequent Collocations of the Two Corpora

As it was described in the previous chapter, this study was conducted through a frequency-based approach in order to identify the most frequent collocations used in Iranian and

Norwegian research articles in sub-disciplines of applied linguistics. Accordingly, a corpus of 132,355 words – consisting of 67,673 words for the Iranian corpus and 64,682 words for the Norwegian corpus – was selected. The two corpora were analysed using *Collocation Extract* software (version 3.06). The following table illustrates the results obtained from the analysis of the two corpora indicating the 15 most frequent collocations.[6]

Table 4.1 Fifteen most frequent collocations in the Iranian and Norwegian corpora

	Iranian corpus			Norwegian corpus		
	Collocation	F	Category	Collocation	F	Category
1	wedding invitation	85	N+N	plural pronouns	69	Adj+N
2	speech act	75	N+N	board members	45	N+N
3	native speakers	70	Adj+N	case study	31	N+N
4	EFL learners	50	N+N	personal pronouns	26	Adj+N
5	discourse community	37	N+N	deafblind person	23	Adj+N
6	lower status	33	Adj+N	sign language	19	N+N
7	research articles*	33	Adj+N	higher education	16	Adj+N
8	present study	29	Adj+N	research articles*	15	Adj+N
9	higher status	26	Adj+N	emergency calls	14	N+N
10	most frequent	23	Adv+Adj	vague reference	12	Adj+N
11	Persian speakers	19	Adj+N	language use*	11	Adj+N
12	applied linguistics	16	Adj+N	different kinds	10	Adj+N
13	equal status	16	Adj+N	present study	10	Adj+N
14	data collection	12	N+N	same time	8	Adj+N
15	personal space	12	Adj+N	study programme	8	N+N
	Total	536		Total	317	

As it can be seen, the 15 most frequent collocations used in the Iranian and Norwegian corpora are illustrated in Table 4.1. According to the above table, the two corpora are not similar in their 15 most frequent collocations; however, there are some similarities. These two sets have two collocations in common, namely 'research articles' and 'present study'. The table shows that 'wedding invitation', with a frequency of 85, was the most frequent collocation used in the Iranian corpus, while the most frequent collocation in the Norwegian corpus was 'plural pronouns' with a frequency of 69. The collocation 'speech act' with a frequency of 75 was the second most frequent among Iranians, and 'native speakers' with a

---

[6] It should be noted that the collocations marked with asterisks in the table above are classified in the Adj+N category even though the first part of the collocation is grammatically a noun. As a result of the adjectival function of the noun, it becomes an attributive noun and considered as an adjective. For detailed descriptions see section 3.4.2. above.

frequency of 70, 'EFL learners' with a frequency of 50, and 'discourse community' with a frequency of 37 were ranked as the third, fourth, and fifth most frequent collocations found in the Iranian corpus. The next five used collocations by Iranians were 'lower status', 'research articles', 'present study', 'higher status', and 'most frequent' with frequencies of 33, 33, 29, 26, and 23, respectively. As it is shown, the frequency range of 11th to 15th used collocations in the Iranian corpus is 19 to 12, which includes 'Persian speakers', 'applied linguistics', 'equal status', 'data collection' and 'personal space' with frequencies of 19, 16, 16, 12, and 12, respectively. Regarding the Norwegian corpus, the five most used collocations were 'plural pronouns', 'board members', 'case study', 'personal pronouns', and 'deafblind person' with frequencies of 69, 45, 31, 26, and 23, respectively. Additionally, the collocation 'sign language' with a frequency of 19 was the sixth most frequent in the Norwegian corpus, followed by 'higher education' with a frequency of 16, 'research article' with a frequency of 15, 'emergency calls' with a frequency of 14 and 'vague reference' with a frequency of 12. The next five collocations were 'language use', 'different kinds', 'present study', 'same time', and 'study programme', which ranked as the eleventh, twelfth, thirteenth, fourteenth, and fifteenth most frequent collocations used in the Norwegian corpus with frequencies of 11, 10, 10, 8, and 8, respectively. The presented data show that the frequency range of the 15 most frequent collocations in the Iranian data was 85 to 12 while this range was between 69 and 8 in the Norwegian data. Also, Iranian scholars used the 15 most frequent collocations 536 times totally, whereas Norwegian scholars used the 15 most frequent collocations found in their corpus 317 times. The difference in frequencies of the 15 most frequent collocations of the two corpora is illustrated in Figure 4.1.
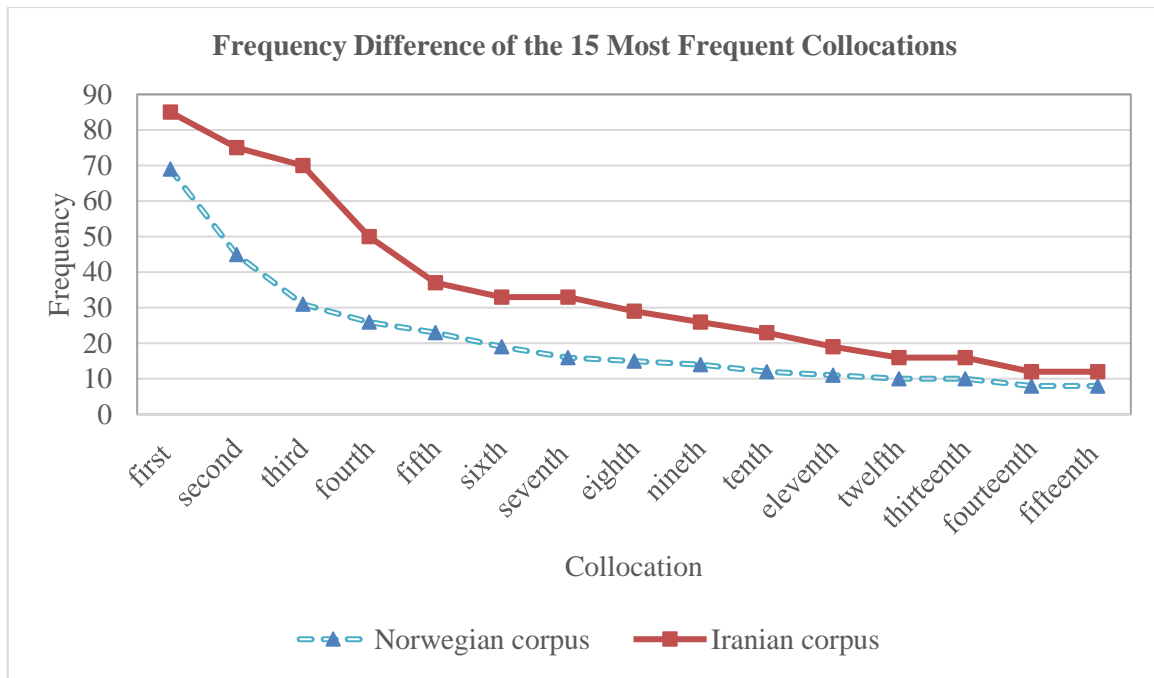
Figure 4.1 Difference in frequencies of the 15 most frequent collocations of the Iranian and Norwegian corpora

Concerning the category of the collocations, three out of six collocation categories can be observed in the table containing the 15 most used collocations of the two corpora. The category of Adj+N is the most frequent in both corpora, followed by the N+N category. Only one collocation in the Iranian corpus was categorized under the Adv+Adj category though this was not observed in the 15 most frequent collocations used by Norwegians. For a better visual understanding of the most used categories, the results are depicted in the following figure.
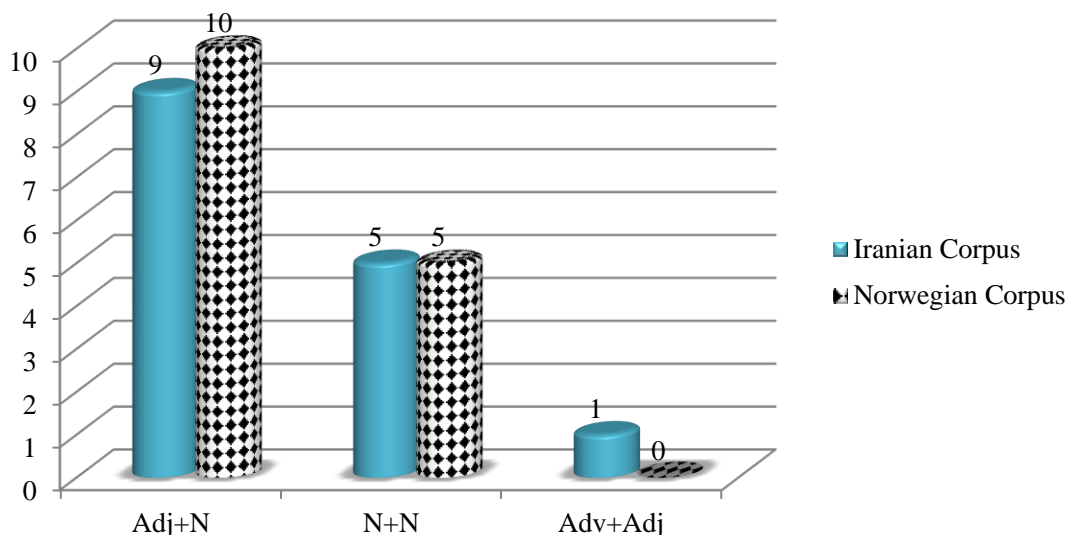
Figure 4.2 Frequency of collocations' categories in 15 most frequent collocations in the two corpora

According to Figure 4.2, 9 out of the 15 most used collocations found in the Iranian corpus were of Adj+N type, while 10 out of the 15 most frequent collocations of the Norwegian corpus were of this type. Moreover, five out of the 15 most frequent collocations of both corpora were categorized as N+N type based on the Oxford Collocations Dictionary. Furthermore, one of the 15 most used collocations found in the Iranian corpus was of Adv+Adj type, which was not observed in the 15 most frequent collocations of the Norwegian corpus.

The entire lists of the frequent collocations in the two corpora are presented in the following sections.

4.3. Frequent Collocations of the Iranian Corpus

The following tables illustrate frequent collocations identified in the Iranian corpus. These collocations are classified based on their associated category checked in the Oxford Collocations Dictionary. Accordingly, frequency and percentage of occurrence of every single collocation are included in the tables.

Table 4.2 Adj+N collocations in the Iranian corpus

	Adjective	F	Noun	F	Adj+N	Percent
1	native	138	speakers	149	70	0.103
2	lower	69	status	195	33	0.048
3	present	52	study	194	29	0.042
4	higher	61	status	195	26	0.038
5	Persian	214	speakers	149	19	0.028
6	applied	27	linguistics	20	16	0.023
7	equal	36	status	195	16	0.023
8	moral	20	order	68	11	0.016
9	personal	32	space	50	11	0.016
10	Iranian	121	society	28	11	0.016
11	following	71	excerpt	24	10	0.014
12	foreign	13	language	174	10	0.014
13	language	174	proficiency	25	10	0.014
14	same	97	time	85	10	0.014
15	American	78	speakers	149	10	0.014
16	language	174	learners	123	10	0.014
17	English	157	speakers	149	10	0.014
18	high	49	level	44	9	0.013
19	same	97	gender	61	9	0.013
20	American	78	counterparts	12	8	0.011
21	Korean	11	speakers	149	8	0.011
22	further	50	research	135	8	0.011
23	academic	41	writers	68	8	0.011
24	verbal	22	communication	60	7	0.01
25	academic	41	writing	40	7	0.01
26	human	22	life	96	7	0.01
27	sexual	17	harassment	9	6	0.008
28	implicit	12	suggestion	29	5	0.007
29	opposite	5	gender	61	5	0.007
30	physical	11	distance	82	5	0.007
31	university	35	campus	6	5	0.007
32	cultural	75	values	19	5	0.007
33	low	17	status	195	5	0.007
	Total				419	0.619

Table 4.2 includes the 33 most frequent Adj+N collocations identified in the Iranian corpus. As it is evident from the table, adjectives are presented in the first column, followed by their frequencies in the second column. In the third column nouns are presented, followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of Adj+N collocations. In the sixth column, the percentage of every single collocation is presented based on the total number of the words in the Iranian corpus. Evidently, 'native speakers' was the most frequent collocation found in the Iranian corpus with 0.103 percent proportion of the total words. The collocation 'lower status' with 0.048

percent proportion was ranked as the second most frequent collocation found in the Iranian corpus. The third one was 'present study', which constituted 0.042 percent of the whole words. The fourth and the fifth ones were 'higher status' and 'Persian speakers' with 0.038 and 0.028 percent proportion of the total words respectively. The next two, ranking as sixth and seventh most frequent Adj+N collocation, were 'applied linguistics' and 'equal status', each of which comprises 0.023 percent of the whole corpus. The next three collocations ranking from 8 to 10 were 'moral order', 'personal space', and 'Iranian society', each including 0.016 percent of the corpus. Each of the next six collocations, ranking from 11 to 17, was used 10 times throughout the corpus constituting 0.014 percent of the whole words in the corpus. These collocations were 'following excerpt', 'foreign language', 'language proficiency', 'same time', 'American speakers', 'language learners', and 'English speakers'. Number 17 and 18 were 'high level' and 'same gender', constituting 0.013 percent of the whole words. Number 19 to 22 were 'American counterparts', 'Korean Speakers', 'further research', and 'academic writers'. Each of these Adj+N collocations was used 8 times throughout the corpus, which accounts for 0.011 percent of the whole words. The collocations 'verbal communication', 'academic writing', and 'human life' were the next three Adj+N collocations, each constituting 0.01 percent of the Iranian corpus. The collocation 'sexual harassment', ranking as 26 in the list, was used 6 times throughout the texts, which means it constituted 0.008 percent of the Iranian corpus. The least used Adj+N collocations found in the Iranian corpus were 'implicit suggestion', 'opposite gender', 'physical distance', 'university campus', 'cultural values' and 'low status', ranking from 27 to 32, each with a frequency of 5 constituted 0.007 percent of the words included in the corpus. As it can be seen, the total number of Adj+N collocations used in the Iranian corpus was 419, which is 0.619 percent of the entire words in the corpus.

In the following table, N+N collocations found in the Iranian corpus are depicted. Just as it was the case in the previous table, the first column is devoted to the first nouns, followed by their frequencies in the second column. In the third column the second nouns are presented, followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of N+N collocations. In the sixth column, the percentage of every single collocation is presented based on the total number of the words in the Iranian corpus.

Table 4.3 N+N collocations in the Iranian corpus

	Noun	F	Noun	F	N+N	Percent
1	wedding	129	invitation	190	85	0.125
2	speech	112	act	122	75	0.110
3	EFL	82	learners	123	50	0.073
4	discourse	151	community	83	37	0.054
5	research	135	articles	52	33	0.048
6	speech	112	community	87	15	0.022
7	data	71	collection	11	11	0.016
8	research	135	question	33	9	0.013
9	discourse	151	analysis	110	9	0.013
10	wedding	129	ceremony	16	7	0.01
11	word	34	choice	16	5	0.007
12	noun	7	phrase	11	5	0.007
13	university	35	students	57	5	0.007
	Total				346	0.505

Based on the results gathered in Table 4.3, 13 different N+N collocations have been identified in the Iranian corpus. As it is evident, 'wedding invitation' with a frequency of 85 and proportion of 0.125 was the most frequent N+N collocation, followed by 'speech act'. 'Speech act' was used 75 times, which constituted 0.11 percent of the words in the Iranian corpus. The third N+N collocation was 'EFL learners' with proportion of 0.073. The collocation 'discourse community' with a frequency of 37 and proportion of 0.054 percent of the words was ranked as the fourth most used N+N collocation throughout the Iranian corpus. The fifth most frequent N+N collocation was 'research articles', including 0.048 percent of the Iranian corpus, followed by 'speech community', including 0.022 percent of the words. 'Data collection' was ranked as the seventh most frequent N+N collocation in the Iranian corpus with 0.016 usage percent. The next two were 'research question' and 'discourse analysis', each with 0.013 percent of usage. Number 10 was 'wedding ceremony', used 7 times, which means it constituted 0.01 percent of the total words in the Iranian corpus. In addition, the least used N+N collocations, ranking from 11 to 13, were 'word choice', 'noun phrase', and 'university students', all with the similar frequency of 5 and proportion of 0.007 percent of the total words in the Iranian corpus. As it is illustrated in Table 4.3, the total number of N+N collocations identified in the Iranian corpus was 346, which constituted 0.505 percent of the total number of the words in this corpus.

The next table indicates the results obtained from the analysis of the corpus searching for V+N collocations. The format of the table is generally similar to the previous one for

N+N collocations. In the first column, identified verbs are gathered, followed by their frequencies in the next column. In the third column nouns are presented, followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of V+N collocations. In the sixth column, the percentage of every single collocation is presented, based on the total number of the words in the Iranian corpus.

Table 4.4 V+N collocations in the Iranian corpus

	Verb	F	Noun	F	V+N	Percent
1	share	18	knowledge	47	7	0.01
2	see	37	table	55	7	0.01
	Total				14	0.02

The preceding table shows that there were two different V+N collocations found in the Iranian corpus. As it can be seen, 'share knowledge' and 'see table', the two collocations of this category shared the same frequency of 7 and proportion of 0.01 percent of the total words of the Iranian corpus. Evidently, this type of lexical collocations was used 14 times throughout the Iranian corpus, which means that it constituted 0.02 percent of the total words in this corpus.

The following table contains Adv+Adj collocations. Accordingly, the first column is devoted to adverbs, followed by their frequencies in the second column. In the third column adjectives are presented, followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of Adv+Adj collocations. In the sixth column, the percentage of every single Adv+Adj collocation is presented, based on the total number of the words in the Iranian corpus.

Table 4.5 Adv+Adj collocations in the Iranian corpus

	Adverb	F	Adjective	F	Adv+Adj	Percent
1	most	95	frequent	34	23	0.033
2	most	95	common	44	11	0.016
	Total				34	0.050

As it is evident from the table, two different Adv+Adj collocations were identified in the Iranian corpus. The collocation 'most frequent' with a frequency of 23 and a proportion of

0.033 percent was the most frequent collocation of this category. The least frequent collocation of this category was 'most common' with a frequency of 11 and a proportion of 0.016 percent of the total collocations. As the table shows, Adv+Adj collocations were used 34 times throughout the Iranian corpus constituting 0.05 percent of the total words of this corpus.

The next table indicates N+V collocations found in the Iranian corpus. So, the first column is devoted to nouns, followed by their frequencies in the second column. In the third column verbs are presented, followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of N+V collocations. In the sixth column, the percentage of every single N+V collocation is presented, based on the total number of the words in the Iranian corpus.

Table 4.6 N+V collocations in the Iranian corpus

	Noun	F	Verb	F	N+V	Percent
1	results	60	indicated	16	5	0.007
2	table	55	shows	27	5	0.007
	Total				10	0.014

According to Table 4.6, only two different N+V collocations were identified in the corpus. These include 'results indicated' and 'table shows', both with a frequency of 5 and a proportion of 0.007 percent of the total collocations. Based on what is indicated in Table 4.6, the total number of N+V collocations is 10 including 0.014 percent of the total number of words in the Iranian corpus.

In the previous tables, usage percentage of the collocations was calculated, based on the total number of the words in the corpus. In order to achieve a better understanding of the usage of different types of lexical collocations in comparison with each other, the percent proportion of each type is calculated according to the total number of lexical collocations identified in the Iranian corpus. Thus, in the following figure, the proportion of each category of collocations is depicted.
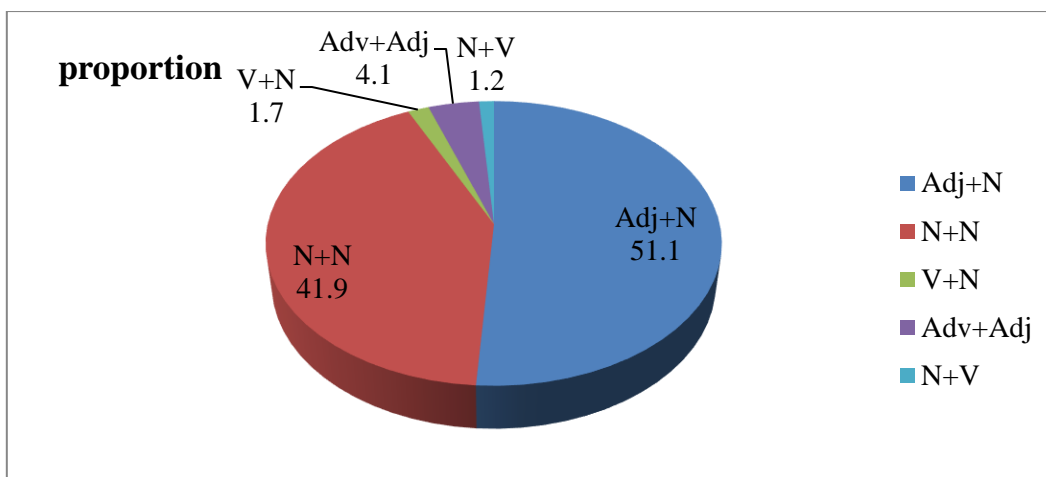
Figure 4.3 Proportion of each type of collocations in the Iranian corpus

According to Figure 4.3, five categories of lexical collocations were identified in the Iranian corpus. Adj+N category of collocations with a proportion of 51.1 percent comprised the highest number of collocations. N+N category of collocations was ranked as the second category with a proportion of 41.9 percent of the total collocations. The third category was Adv+Adj with a proportion of 4.1 percent of the total collocations, followed by V+N category with a proportion of 1.7 percent as the fourth one. The least used category with the lowest number of identified collocations was N+V with a proportion of 1.2 percent of the total collocations.

4.4.   Frequent Collocations of the Norwegian Corpus

Similar to the previous section, the following tables illustrate frequent collocations identified in the Norwegian corpus. These collocations are classified based on their associated category in the Oxford Collocations Dictionary. Accordingly, frequency and percent of occurrence of all the collocations are included in the tables.

Table 4.7 includes Adj+N collocations found in the Norwegian corpus. Accordingly, the first column includes identified adjectives; the second column depicts the adjectives' frequencies; the third one includes nouns, followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of Adj+N collocations. And finally, in the sixth column, the percentage of all the Adj+N collocations is presented, based on the total number of the words in the Norwegian corpus.

Table 4.7 Adj+N collocations in the Norwegian corpus

	Adjective	F	Noun	F	Adj+N	Percent
1	plural	99	pronouns	205	69	0.106
2	personal	85	pronouns	153	26	0.040
3	deafblind	56	person	83	23	0.035
4	higher	32	education	23	16	0.024
5	language	208	use	96	11	0.017
6	vague	31	reference	43	11	0.017
7	present	97	study	118	10	0.015
8	different	131	kinds	22	10	0.015
9	same	48	time	60	8	0.012
10	Nordic	7	countries	24	7	0.010
11	social	69	implications	16	7	0.010
12	foreign	18	students	86	7	0.010
13	singular	14	pronouns	153	7	0.010
14	mother	10	tongue	5	5	0.007
15	European	33	integration	8	5	0.007
16	science	15	programme	31	5	0.007
17	emotional	17	expressions	30	5	0.007
18	native	17	speakers	31	5	0.007
19	medical	27	texts	28	5	0.007
20	national	29	culture	27	5	0.007
21	previous	31	studies	37	5	0.007
22	official	8	language	208	5	0.007
23	definite	48	article	41	5	0.007
24	foreign	18	language	208	5	0.007
25	different	131	forms	32	5	0.007
26	national	29	language	208	5	0.007
	Total				277	0.412

Table 4.7 includes word combinations which are classified as Adj+N collocations. As it can be seen, 27 combinations are identified as Adj+N collocations in the Norwegian corpus. Obviously, 'plural pronouns' with a frequency of 69 and a proportion of 0.106, is the most frequent Adj+N collocation of the Norwegian corpus, followed by 'personal pronouns' with a frequency of 26 and a proportion of 0.040 percent of the total words in the corpus. The collocation 'deafblind person' with a frequency of 23 and a proportion of 0.035 percent of the total words in the corpus is placed third in the list. The fourth one was 'higher education', which constituted 0.024 percent of the total words of the corpus. Number five and six in the list are 'language use' and 'vague reference' with equal frequencies of 11 and usage ratio of 0.017 percent. The next two Adj+N collocations of the Norwegian corpus are 'present study' and 'different kind', each with a frequency of 10, which constituted 0.015 percent of the corpus. 'Same time' with a frequency of 8 and a 0.012 usage percentage occupied the ninth

rank of the table. Numbers 10 to 13, with similar frequencies of 7 and a usage percentage of 0.010, were 'Nordic countries', 'social implications', 'foreign students', and 'singular pronouns'. Finally, the collocations 'mother tongue', 'European integrations', 'science programme', 'emotional expressions', 'native speakers', 'medical texts', 'national culture', 'previous studies', 'official language', 'definite article', 'foreign language', 'different forms', and 'national language', with a frequency of 5 and a proportion of 0.007 percent were identified as the least frequent collocations of this category. As it is evident, the total number of Adj+N collocations used in the Norwegian corpus was 277, which compromised 0.412 percent of the total number of the words in the corpus.

The next table indicates the results obtained from the analysis of the corpus searching for N+N collocations. Similar to the previous table, the first column includes the first nouns of the combinations, followed by their frequencies in the second column. In the third column, the second nouns of the combinations are listed, which are followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of N+N collocations. In the sixth column, the percentage of each collocation is presented, based on the total number of the words in the Norwegian corpus.

Table 4.8 N+N collocations in the Norwegian corpus

	Noun	F	Noun	F	N+N	Percent
1	board	63	members	74	45	0.069
2	case	95	study	118	31	0.047
3	sign	36	language	208	19	0.029
4	research	66	articles	71	15	0.023
5	emergency	32	calls	42	14	0.021
6	study	118	programme	31	8	0.012
7	seating	8	arrangement	8	7	0.010
8	health	11	care	12	7	0.010
9	board	63	meeting	56	6	0.009
10	language	208	learning	19	5	0.007
	Total				157	0.242

Table 4.8 illustrates the collocations which are classified as N+N collocations. According to the table above, ten collocations are identified as N+N collocations in the Norwegian corpus. As it is observable, 'board members', with a frequency of 45 and a proportion of 0.069 was the most frequent collocation of the Norwegian corpus. The second most frequent N+N

collocation was 'case study' with a frequency of 31 and a proportion of 0.047 percent of the total number of the words in the corpus. The collocation 'sign language', with a frequency of 19 and usage percentage of 0.029, occupied the third rank in the list. This is followed by 'research articles' and 'emergency calls', which constituted 0.023 and 0.021 percent of the total words of the corpus respectively. The sixth frequent N+N collocation of the Norwegian corpus was 'study programme', which was used 8 times, constituting 0.012 percent of the corpus. Number 7 and 8 of the corpus were 'seating arrangement' and 'health care' with equal frequencies of 7 and usage percent of 0.010. The next one was 'board meeting' with 0.009 percent of usage occupying the ninth rank in the table. Finally, the least used N+N collocation of the corpus was 'language learning', which was used five times throughout the texts and constituted 0.007 percent of the Norwegian corpus. Based on what is indicated in Table 4.8, the total number of N+N collocations of the Norwegian corpus was 157, which constituted 0.242 percent of the entire corpus.

The following table illustrates the results obtained from the analysis of the corpus searching for V+N collocations. Thus, the first column includes verbs of the combinations, followed by their frequencies in the second column. In the third column, nouns of the combinations are presented, followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of V+N collocations. In the sixth column, the percentage of each collocation is presented, based on the total number of the words in the Norwegian corpus.

Table 4.9 V+N collocations in the Norwegian corpus

	Verb	F	Noun	F	V+N	Percent
1	make	46	sense	47	7	0.010
2	provide	30	information	104	6	0.009
3	shed	5	light	7	5	0.007
4	assign	17	tasks	16	5	0.007
	Total				23	0.003

As it is indicated in the above table, four collocations were identified as frequent V+N collocations in the Norwegian corpus. The collocation 'make sense' was the most frequent collocation with a frequency of 47 and a proportion of 0.010 percent of the total words in the Norwegian corpus. The second collocation was 'provide information' with a frequency of 6

and a proportion of 0.009 percent. The third and fourth ones were 'shed light' and 'assign tasks' with equal frequencies of 5 and proportions of 0.007 percent. According to the data in Table 4.9, this type of collocations was used 23 times throughout the Norwegian corpus, which included 0.003 percent of the words in the text.

Table 4.10 contains Adv+Adj collocations identified in the Norwegian corpus. Accordingly, the first column includes identified adverbs; the second column depicts the adverbs' frequencies; the third one includes adjectives, followed by their frequencies in the fourth column. The fifth column is devoted to the frequencies of co-occurrences of Adv+Adj collocations. And finally, in the sixth column, the percentage of every single Adv+Adj collocation is presented based on the total number of the words in the Norwegian corpus.

Table 4.10 Adv+Adj collocations in the Norwegian corpus

	Adverb	F	Adjective	F	Adv+Adj	Percent
1	most	57	likely	21	5	0.007

According to the table above, only one collocation was assigned to this category of collocations. As it is evident, 'most likely' with a frequency of 5 and a proportion of 0.007 was identified as the only Adv+Adj collocation in Norwegian corpus.

In the previous tables, the usage percentage of the collocations was calculated based on the total number of the words in the Norwegian corpus. In order to obtain a better understanding of the usage of different types of lexical collocations in comparison with each other, the percent proportion of each type is calculated according to the total number of lexical collocations identified in the Norwegian corpus. Thus, in the following figure, the proportion of each category of collocations is depicted.
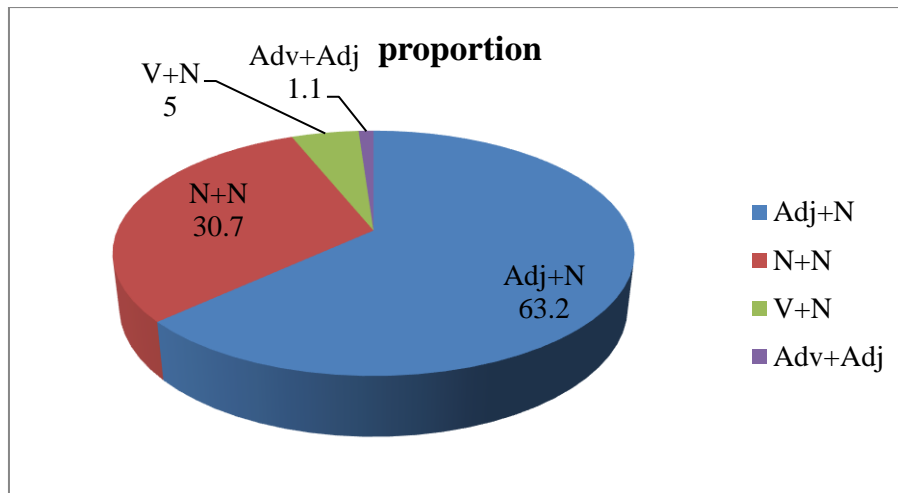
Figure 4.4 Proportion of each type of collocations in the Norwegian corpus

Figure 4.4 illustrates that four types of lexical collocations were identified in the Norwegian corpus. Based on what is depicted in the figure, 63.2 percent of the collocations were of Adj+N type; 30.7 percent were categorized as N+N type; 5 percent were identified as V+N type; and 1.1 percent were categorized as Adv+Adj type.

4.5. Common Collocations of the Two Corpora

In the following table, common collocations found in the two corpora are presented. The first column indicates the common collocations of the two corpora, followed by their types in the second column. The third and fourth columns show the frequencies of the collocations in the Iranian and Norwegian corpora respectively.

Table 4.11 Common collocations of the two corpora

ID	Common collocations	Types of collocations	Frequency in the Iranian corpus	Frequency in the Norwegian corpus
1	research articles	N+N	33	15
2	present study	Adj+N	29	10
3	foreign language	Adj+N	10	5

According to Table 4.15, there were only three collocations in common in the Iranian and the Norwegian corpora. As it can be seen, one out of the three common collocations is of N+N type while the other two are of Adj+N type. The collocation 'research articles', with a frequency of 33 in the Iranian corpus and 15 in the Norwegian corpus, was the most frequent among the three common collocations in both corpora. The second one is 'present study' with a frequency of 29 in the Iranian corpus and 10 in the Norwegian corpus. And the third one, which was the least used one among the three, was 'foreign language' with a frequency of 10 in the Iranian corpus and 5 in the Norwegian corpus. The following figure illustrates the percent proportion of the common collocations comparing to the total number of collocations used in each corpus.
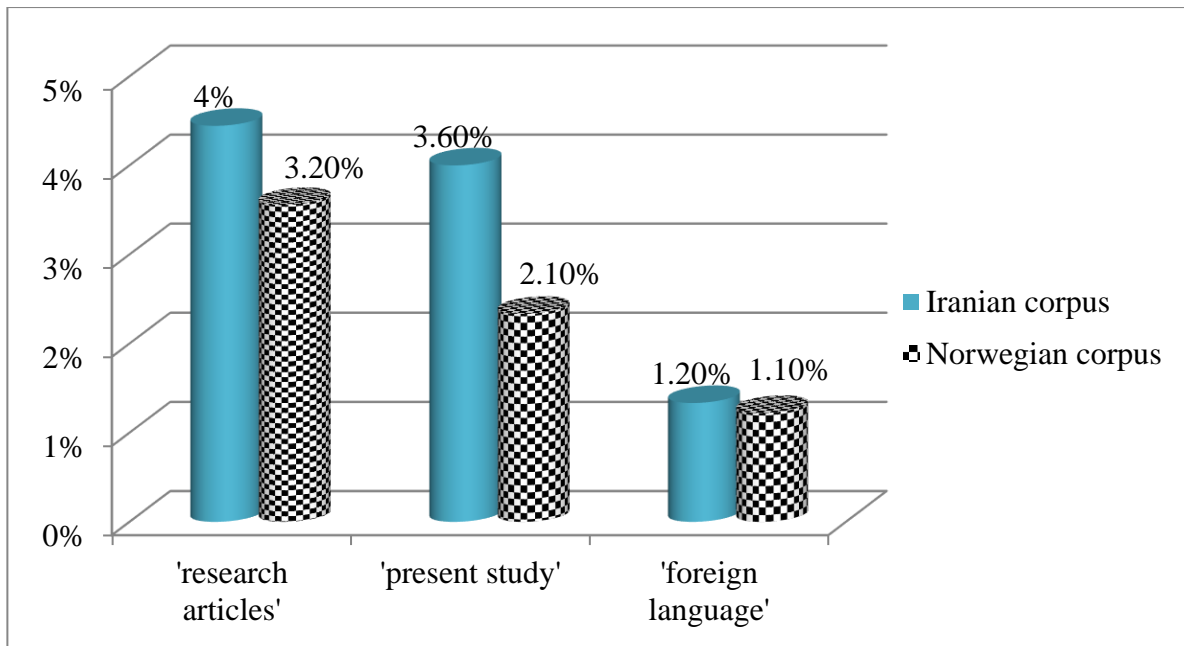
Figure 4.5 Comparison of common collocations

As it is illustrated in Figure 4.5, the collocation 'research articles' constituted 4% of the total number of collocations identified in the Iranian corpus and 3.2% of collocations identified in the Norwegian corpus. The collocation 'present study' constituted 3.6% of the total number of collocations of the Iranian corpus while it constituted 2.1% of the collocations of the Norwegian corpus. The collocation 'foreign language' constituted 1.2% and 1.1% of the total number of collocations found in the Iranian and Norwegian corpora respectively. As it can be seen, the differences between the proportions of all three collocations were not very much; though the difference between the proportions of the collocation 'present study' was higher than the two others.

4.6.   Comparing Collocation Types of the Two Corpora

The following tables compare the frequency of collocations in each category of the two corpora. Since the frequency occurrence in a corpus is largely corpus-size dependent and does not necessarily reveal whether the occurrence is significantly frequent or the difference between the results is significant, the data should be normalized before performing any inferential statistics. To obtain the desired results, frequencies of collocations of each

category were normalized per 1,000,000 words. Then, a Chi-square test was employed to see whether there were significant differences in the use of collocations between the two corpora.

Table 4.12 Chi-square test for the differences in Adj+N type of collocations between the two corpora

Adj+N	Value	df*	Asymp. Sig.** (2-sided)
Pearson Chi-Square	59.000	21	0.000
Likelihood Ratio	80.959	21	0.000
Linear-by-Linear Association	0.206	1	0.650
No. of Valid Cases	59		

Key:
*df: degree of freedom
**Asymp. Sig.: Asymptotic Significance

As it can be seen, the Adj+N category of the Iranian corpus was compared to the same category belonging to the Norwegian corpus through Pearson Chi-square test. There are different items in the table above but *Asymp. Sig.* column is the one which directly helps us answer the question and accept or reject the null-hypothesis. In comparing variables, when *Asymp. Sig.* (or ρ *value*) is $\geq$ 0.05 it means that there is no significant difference between the observed variables. On the other hand, if *Asymp. Sig.* is $\leq$ 0.05 it means that there is a significant difference between the observed variables. According to the table above, the result of the Chi-square test shows that there was a significant difference (sig = 0.000) in the use of collocations of this category between the two corpora.

Table 4.13 Chi-square test for the differences in N+N type of collocations between the two corpora

N+N	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	46.000	36	0.123
Likelihood Ratio	42.655	36	0.207
Linear-by-Linear Association	6.230	1	0.013
No. of Valid Cases	23		

Table 4.12 compares the N+N categories of the two corpora. The results of Pearson Chi-square test revealed an Asymp. Sig. value of 0.123. Since the obtained value should be ≤ 0.05 for a significant difference, it can be said that there was no significant difference in the use of this type of collocations between the two corpora.

Table 4.14 Chi-square test for the differences in V+N type of collocations between the two corpora

V+N	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.000	3	0.112
Likelihood Ratio	7.638	3	0.054
Linear-by-Linear Association	1.495	1	0.221
No. of Valid Cases	6		

In the above table, the results of comparing V+N categories of the two corpora are depicted. Performing Pearson Chi-square test for the V+N category showed that there was no significant difference between the two corpora in the use of V+N collocations. This is due to the p value (sig = 0.112), which is greater than the accepted level of 0.05. If it was ≤ 0.05, the

two groups of the writers were different in using V+N collocations. But, now that the results show the opposite, it can be concluded that they were not different in using this type of lexical collocations in their academic articles.

Table 4.15 Chi-square test for the differences in Adv+Adj type of collocations between the two corpora

Adv+Adj	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	6.000	4	0.199
Likelihood Ratio	6.592	4	0.159
Linear-by-Linear Association	2.000	1	0.157
No. of Valid Cases	3		

In Table 4.14, Adv+Adj category of lexical collocations found in the two corpora is compared. The result of the comparison between the Adv+Adj category of the two corpora showed that there was no significant difference. As it can be seen, the obtained Asymp. Sig. is 0.199 whereas the accepted level depicting a significant difference is ≤ 0.05. Since, this value is greater than the accepted level, it should be said that the two groups of the scholars were not significantly different in using Adv+Adj collocations.

5. Discussion

5.1. Overview

This chapter provides a summary of the findings of the study. The summary is followed by the discussion of the findings in form of answers to the first and second research questions of the study.

5.2. Summary of the Findings

The findings of the present study are summarized as follows:

The Iranian corpus:

1. A total of 818 collocations were identified in the Iranian corpus through a frequency-based analysis method.
2. Out of 818 identified collocations of the Iranian corpus, 419 occurrences were classified as frequent Adj+N collocations in 33 different combinations as presented in Table 4.2 above.
3. Out of 818 identified collocations of the Iranian corpus, 346 occurrences were classified as frequent N+N collocations in 13 different combinations as presented in Table 4.3 above.
4. Out of 818 identified collocations of the Iranian corpus, 14 occurrences were classified as frequent V+N collocations in 2 different combinations as presented in Table 4.4 above.
5. Out of 818 identified collocations of the Iranian corpus, 34 occurrences were classified as frequent Adv+Adj collocations in 2 different combinations as presented in Table 4.5 above.
6. Out of 818 identified collocations of the Iranian corpus, 10 occurrences were classified as frequent N+V collocations in 2 different combinations as presented in Table 4.6 above.

7. A total of 462 collocations were identified in the Norwegian corpus through a frequency-based analysis method.

8. Out of 462 identified collocations of the Norwegian corpus, 277 occurrences were classified as frequent Adj+N collocations in 26 different combinations as presented in Table 4.7 above.

9. Out of 462 identified collocations of the Norwegian corpus, 157 occurrences were classified as frequent N+N collocations in 10 different combinations as presented in Table 4.8 above.

10. Out of 462 identified collocations of the Norwegian corpus, 23 occurrences were classified as frequent V+N collocations in 4 different combinations as presented in Table 4.9 above.

11. Out of 462 identified collocations of the Norwegian corpus, 5 occurrences were classified as frequent Adv+Adj collocations in only one combination as presented in Table 4.10 above.

Common Findings in the two corpora:

12. Adj+N category of collocations was the most frequent in both the Iranian and the Norwegian corpora.

13. N+V category of collocations was the least frequent in the Iranian corpus, while Adv+Adj was the least frequent category in the Norwegian corpus.

14. No matches were found for N+V collocations in the Norwegian corpus. Also, Adv+V was not found in any of the corpora.

15. Analysis of the two corpora to identify common collocations resulted in a total of 3 collocations.

16. Out of the 3 common collocations, 2 belonged to the Adj+N category and one belonged to the N+N category.

17. The results of Pearson Chi-square test revealed that there was a significant difference in the frequency of Adj+N collocations between the Iranian and the Norwegian corpora while no difference was observed in the use of N+N, V+N, and Adv+Adj types of collocations.

5.3. Discussion of the Findings

The main discussion of this study which deals with answering the research questions and discussing the findings of the study with regard to the main scopes of the study, namely corpus linguistics and collocations is presented in two phases. In the first phase, the answer to the first question of the study, 'What are the most frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars?' is discussed. In the second phase, the answer to the second research question, 'Is there any significant difference between the frequent lexical collocations in research articles of applied linguistics written in English by Iranian and Norwegian scholars?' is discussed.

5.3.1. Answer to the First Research Question

Taking the first research question into account, this phase of discussion deals with the relevant aspects derived from the analysis of the results presented in the previous chapter. As it was stated earlier, collocations are mainly divided into two categories namely grammatical collocations and lexical collocations. Grammatical collocations typically consist of a noun, a verb, or an adjective, plus a grammatical particle such as a preposition, an infinitive or a clause. Unlike grammatical collocations, lexical collocations do not include grammatical portions; they include different combinations of nouns, verbs, adjectives, and adverbs. These combinations are Verb + Noun, Adjective + Noun, Noun + Verb, Noun + Noun, Adverb + Adjective, and Verb + Adverb. Since the first question of this study investigates lexical collocations used in the corpora, searching for grammatical collocations are neglected and the concentration is only on the above-mentioned combinations of lexical collocations.

To answer this question, a total of 132,355 words including two corpora of research articles written by Iranian and Norwegian scholars were analyzed. Analysis of the data was conducted using *Collocation Extract* software to identify the most frequent collocations used in the two corpora through a frequency-based approach.

Analyzing the data revealed that a total of 818 lexical collocations were used in the Iranian corpus. Also, a total of 462 lexical collocations were identified in the Norwegian corpus. According to the obtained results, Adj+N and N+N collocations were the most frequent types of collocations identified in both corpora. That is writers of both groups were more inclined to use these two types of lexical collocations though common collocations

were rare. Only three common collocations were identified in both corpora, out of which two belonged to Adj+N type of collocations and one belonged to N+N category; however, all three were among the most frequent collocations used in both corpora. This interesting result addressed two opposing but related points. On the one hand, one could consider these few common collocations due to the different sub-disciplines that require different collocational combinations. On the other hand, it should not be forgotten that despite differences in sub-disciplines, all the articles were taken from the *Journal of Pragmatics* which implies a common discipline in the field of applied linguistics. Thus, the few number of common collocations is quite surprising.

Concerning the most frequent types of lexical collocations in the Iranian corpus, as it was mentioned earlier, Adj+N type of collocations were the most frequent ones, followed by N+N type of collocations. However, the collocation 'wedding invitation', categorized as N+N, was the most used collocation identified in the Iranian corpus. This was followed by another N+N collocation, 'speech act', ranking as the second most frequent collocation in the Iranian corpus. The next one was 'native speaker', occupying the third rank of the frequent collocations. Again, the following two collocations were of N+N type, namely 'EFL learners' and 'discourse community'. This indicates that although Adj+N was the most frequent type of collocations identified in the Iranian corpus, four out of the five top most used collocations were of N+N type.

On the contrary, the least used type of collocation in the Iranian corpus was N+V category. Two collocations were identified for this category, namely 'results indicated' and 'table shows', each with a minimum percentage of 0.007 in the whole Iranian corpus. This finding is slightly different from Bahardoust (2012: 185-200), who found that Verb+Noun and Adjective+Noun collocations were used the most and Noun+Verb collocations were used the least by Iranian EFL students.

However, there were other collocations with minimum usage percent. As it was revealed by the results, other collocations such as 'implicit suggestion', and 'word choice', categorized as Adj+N and N+N, had the minimum usage percent. On the other hand, other collocations were identified with moderate usage such as 'English speakers', 'data collection', 'share knowledge', and 'most common'. Such collocations were classified under different categories of collocations; this implies that they are frequently used by Iranian scholars in their writings. This is somehow unlike what was revealed by HassanAbadi (2003:

45-59), who found that Iranian EFL learners learnt V+N category better than other categories of lexical collocations. If V+N collocations were easier to learn, then the number of this specific type of collocation should have been much higher than what was found in the Iranian corpus.

Regarding frequent types of lexical collocations in the Norwegian corpus, the results showed that, similar to the Iranian corpus, Adj+N type of collocations was the most frequent one, followed by N+N type. This is quite different from the findings of El-Dakhs (2015: 60-74). She found that V+N type of lexical collocations was the most frequent type of collocations used by the participants. However, it should be noted that the participants of her study were Arab students; so both their level of English competence and first language may affect the results. Though Shokouhi and Mirsalari (2010: 1-24) rejected the effect of linguistic knowledge on the collocational knowledge of the learners. Also, Hassanabadi (2003: 45-59) found that V+N type of collocation was easier to learn, which is again different from the findings of this study.

Unlike the results obtained from the Iranian corpus, the most frequent collocation of the Norwegian corpus was 'plural pronoun', which is an Adj+N type of collocation. This was followed by 'board members', as the second most frequent collocation in the Norwegian corpus, which was categorized as a N+N type of collocation. The next one, 'case study' was a N+N collocation, which was followed by two Adj+N collocations. Accordingly, in the Norwegian corpora, three out of five most frequent collocations were in the Adj+N category.

Concerning the least frequent type of collocations identified in the Norwegian corpus, it was found that Adv+Adj was the least frequent type of collocation found in the Norwegian corpus. Only one collocation, 'most likely', was categorized as Adv+Adj type of collocation with minimum usage percent of 0.007. However, 'most likely' was not the only collocation with the least frequency. There were other collocations such as 'mother tongue', 'language learning', 'make sense', etc. which were categorized as Adj+N, N+N, and V+N types of collocations. Moreover, it should be kept in mind that between the continuum of most and least frequent collocations, there were other collocations with moderate usage categorized as Adj+N, N+N, and V+N types of collocation. Meanwhile, it should be noted that no collocation of N+V type was found in the Norwegian corpus. The above-mentioned findings are somehow in contrast to the results of the research by Bahardoust (2012: 185-200), who investigated the rate of lexical collocations in Iranian EFL learners' writings and found that

the rates of V+N and Adj+N collocations were the highest while the rate of N+V collocations was the lowest one. As it was revealed in this study, Adj+N and N+N collocations had the highest rate.

As it was seen, most of the identified collocations belonged to the Adj+N and N+N collocations type and contribution of other types was small. This could be due to different intervening factors such as L1, L2, lack of proper teaching, etc. By looking at the current collocational dictionaries, it is evident that there are thousands of collocations in English which native speakers acquire and use unconsciously but non-native speakers should learn them through different methods. For example, *Collins COBUILD English Words in Use* includes about 100,000 collocational examples or *The BBZ Combinatory Dictionary of English* offers over 70,000 combinations. Obviously, learning all of these combinations and phrases is a very hard task if not impossible. Thus, non-native speakers normally learn and use those collocations to which they are exposed or those which are related to their fields of study. In other words, they learn and use collocations in very restricted contexts. As Tajalli (1994: 124) mentioned, exposure or lack of exposure to a certain type of collocation could affect learning of that specific type. This could be a good explanation of why some types of collocations such as Adj+N are used more frequently than other types.

On the other hand, not all the collocations to which non-natives are exposed are correctly learnt and used. Learning collocations is not easy since there are different obstacles in the process of learning. One of these problems is lack of equivalence for the collocations in the mother tongue. As Marton (1977) stated:

> But the notion of conventional syntagms finds its full dimension only when it is considered contrastively, i.e. when the native language of the learner and his target language are taken into consideration. Real learning problems are caused by these syntagms which are, comparatively speaking, lexically non-congruent, or, in other words, by those in which there is no direct translational equivalence between their corresponding elements.
>
> (Marton 1977: 40-4l)

Surprisingly, this seems true when we look more meticulously at the collocations obtained from the Iranian corpus. The obtained collocations from the Iranian corpus can be categorized under two main groups based on their learning methods. Some are those which have clear direct translational equivalence in Persian and others are those which are frequently used in

the specialized texts of applied linguistics which means scholars are continuously exposed to. For example, 'present study' as one of the frequent collocations illustrated by the results obtained from the Iranian corpus is a collocation that has direct translational equivalence in Persian (i.e., "مطالعه حاضر"). On the other hand, 'EFL learners' is of those collocations that is found frequently in applied linguistics texts. This is also proved by Gitsaki (1996: 7), who found that more frequently-used collocation types can be learnt easier.

But another factor which should not be ignored concerns errors. Occasionally, writers use collocations but not properly, which results in errors and these are not identified by the software. Darvishi (2011: 52-56) identified different sources of errors such as interference of mother tongue, lack of the collocational concept, the interlingual or intralingual transfer, paraphrase and shortage of collocational knowledge. These errors normally result in lack of proper use of collocations. Although detecting the errors was not among the aims of this study, it should be mentioned that some combinations were used more frequently in the corpora. It is probable that the authors tend to use them as collocations but in fact they were collocational errors.

Moreover, it should be noted that lexical collocations are not limited to these five types of collocations. Since the purpose of this study was to identify the frequent two-word strings of collocations in the Iranian and the Norwegian corpora using the *Collocation Extract* software, no other word combination including 3- or more word collocation was intended to be identified.

5.3.2. Answer to the Second Research Question

To answer the second research question, first, frequencies of collocation types were normalized per 1,000,000 words, then, inferential statistics in terms of Chi-square test was employed using *SPSS* version 22.0. Chi-square test is the proper test for comparing frequencies of words between the two corpora because by using this type of test, it can be calculated whether the observed difference between the groups is statistically significant or due to chance (Lindquist 2009 :38).

According to the results obtained from the comparison of frequencies of Adj+N collocations, a significant difference was observed in this category between the two corpora. This indicates that usage of these type of collocations was not similar in the Iranian and the

Norwegian corpora. According to the frequencies of collocations, the obtained results revealed that Iranian scholars used this type of collocations more than their Norwegian counterparts. On the other hand, the two corpora only had two collocations of this type in common, which implies that not only the two corpora had a significant difference in their number of identified Adj+N collocations, but also that they had a low number of collocations in common. Use of a large number of collocations does not entail appropriate use of collocations; sometimes, non-native users overuse these expressions (Chen and Baker 2010: 30-49). González and Ramos (2013: 563-570) had similar findings in their research. Their results showed that despite the use of collocations by learners, their choices lacked the variety, sophistication, and correctness exhibited by native speakers. Therefore, although the two corpora were different in the use of this type of collocations, one cannot figure out if these collocations were overused or used appropriately unless further research would be conducted with this particular purpose.

Regarding N+N collocations, the two corpora had only one collocation in common. Based on the results obtained from the Chi-square test, no significant difference was observed in the use of this category. This means that the tendency to use this type of collocations by both Iranian and Norwegian scholars was similar in the two corpora regarding the number of usage. However, similar to Adj+N category of collocations, N+N collocations had little similarity in terms of the common collocations between the two corpora.

Concerning V+N collocations, the Iranian and the Norwegian corpora had no collocations in common. But the results of the Chi-square test for this category revealed no significant difference in the frequencies of V+N collocations between the two corpora. This suggests that despite the fact that the used collocations were different, the number of usage of this type of collocations were similar in the two corpora.

Taking the results of the comparison of Adv+N category into account, the findings showed that the two corpora had no collocations in common. On the other hand, the results of the Chi-square test revealed no significant difference in the frequency of this type of collocation between the two corpora. Accordingly, it can be said that Iranian and Norwegian scholars had similar tendencies in using this type of collocations though their choice of collocations differed.

Finally, in case of N+V type of collocations, the obtained results revealed that only Iranian scholars used this type of collocations. All in all, from the discussion of this section it

can be concluded that the Iranian and the Norwegian corpora had similarities in usage of N+N, V+N, and Adv+Adj types of collocations, though they had a few collocations in common. On the other hand, they were different in usage of Adj+N collocations. This implies that Iranian scholars used more Adj+N collocations than their Norwegian counterparts. Having a few collocations in common or significant differences in the use of collocations could perhaps be due to the effect of L1 on L2 production as Fan (2009: 110-123) found that Hong Kong ESL students' use of collocations was not only affected by their L1, but it was also affected by their L2 as well as their inadequacy in the lexis and grammar of the target language.

Moreover, few common collocations could be due to the sub-disciplines in which the collocations were used. Although both corpora were compiled of the research articles written by scholars in the field of applied linguistics published in the *Journal of Pragmatics*, the sub-disciplines were different to some extent. Since every discipline has its own terminology, this could be extended to specific collocations for each discipline. For example, 'EFL learner' is a collocation that is mostly used in the sub-discipline of TEFL. Accordingly, this would result in the production of collocations that are specific to a certain sub-discipline. This is what Leech (2001: 3-4) called coverage of register, referring to "the extent to which a word is likely to occur in different varieties of the language". Since the two corpora of this research included articles written in the field of applied linguistics and were not limited to two or three sub-disciplines, the very collocations were not similar, except for some cases.

The above-mentioned point highlights the fact that similar words have their own specific collocations and produce different combinations across the disciplines and written academic genres. Evidently, each genre requires its specific terminology and conventions that each member should use and follow in order to be recognized as a member of that genre's community. Therefore, it is expected that the academic writers be aware of those features and use them professionally. This needs the teachers endeavor in raising learners' awareness to the fact that knowledge is socially constructed within particular domains and this thinking style should be reflected in their writing. Hyland (2008: 561) proposed that teachers should consider text as a dynamic 'social interaction' and teach learners through a 'genre approach' rather than teaching language as a grammatical string of words. Teachers at advanced writing or ESP classes can raise the learners' awareness to these features of their second or foreign language and help them become professional writers in their specific genres.

As it was the case with the present study, by looking at the tables 4.1 to 4.10, one could somehow guess the corpora's related field of study which implies that the writers of both corpora were successful in using specific collocations of their field of study. However, which group of the writers were more successful, or which needs more awareness and training are the questions which cannot be answered by the presented results and require further analysis.

6. Conclusion

The final chapter of the current thesis provides the conclusion. The theoretical and practical implications as well as the limitations of the study are presented next. At the end, some recommendations for further research are put forward.

The present study was a preliminary attempt to identify the most frequent collocations used by Iranian and Norwegian scholars in sub-disciplines of applied linguistics. To this end, a corpus of research articles written by Iranian and Norwegian scholars was compiled. To compile the data, nine research articles written by Iranian scholars in sub-disciplines of applied linguistics and eight research articles written by Norwegian scholars in sub-disciplines of applied linguistics were selected. All the articles were published in the Journal of Pragmatics. The corpus included a total of 132,355 words – consisting of 67,673 words for the Iranian corpus and 64,682 words for the Norwegian corpus.

Then, both the Iranian and the Norwegian corpora were analyzed using a frequency-based approach to identify the most frequent collocations and to check whether there is any probable difference in the use of collocations between the two corpora. In doing so, the data was analyzed using *Collocation Extract* software. Afterwards, the output was analyzed through descriptive and inferential statistical methods using *SPSS* version 22.0.

On the basis of the obtained results, it was revealed that the Adj+N category, with the highest number of collocations, was the main type of collocations used by both Iranian and Norwegian scholars in their research articles. Moreover, it was found that Iranian scholars used more collocations of this type in comparison to their Norwegian counterparts, though this was not the case with other categories. The N+N category was the second most frequent type of collocations identified in both corpora with no significant difference in the usage of this type of collocations between the two corpora. Therefore, it can be concluded that both groups used this type of collocations equally. Concerning other types of collocations such as N+V, Adv+Adj and V+N, the results revealed a low level of usage in both corpora. This could be the effect of L1 on the production of L2, though it needs further research to be confirmed or rejected. Thus, it can be said that perhaps both groups of scholars may not be very familiar with these types of collocations and their usage in academic writing that has resulted in a small number of collocations of these types. This could signal the necessity of

teaching collocations and their appropriate usage as well as raising the awareness of research article writers of different types of collocations and appropriate usage of them to enhance the quality of the academic writings. But it should be born in mind that teaching collocations is not an easy task. As Marton (1977: 43) argued that advanced learners cannot acquire knowledge of conventional syntagms or collocations by mere exposure to the target language. He noted that teachers should pay attention to learners' effective learning of these aspects of the target language (1977: 54). This could probably be reached if, as Martyńska (2004: 2-12) stated, words are learnt in chunks which are collocations to achieve native-like fluency.

Besides, some types of collocations are easier to learn and some are more difficult. As it was revealed in this study, N+N and Adj+N types of collocations were used more than other types in both corpora which shows that both the Iranian and Norwegian scholars have learned these types of collocations better than other types. Reason for this may lay in the fact that they were exposed to these types of collocations more than other types. Although grammatical collocations were not included in the scope of this study, it should be reminded that since learning lexical collocations are easier than the grammatical ones (Shokouhi and Mirsalari 2010: 14), further attention should be paid to teaching and learning grammatical collocations.

Another point which should be addressed here is the number of common collocations used by both groups of the scholars. According to the results, number of common collocations was very few. As it was discussed in chapter five, it could be due to different genres and sub-disciplines in which the articles were written in, since each genre and sub-discipline may require its own specialized vocabulary and phraseology. But the arguing point is that although the sub-disciplines may differ, they all were taken from the *Journal of Pragmatics*, which implies 'pragmatics' as the main discipline of both corpora. Accordingly, it could be argued that since all the articles of both the Iranian and Norwegian corpora were related to pragmatics, more common collocations were expected to be found. The reason for this expectation lies under the fact that readers normally review the academic writings through skimming and scanning to see whether they are useful for their purpose. Since collocations comprise a large portion of the specialized vocabularies (i.e., keywords) of a particular discipline, the readers normally skim and scan across the text looking for these combinations. Accordingly, when the articles are related to a specific discipline, the number of common keywords is expected to be slightly high.

On the other hand, the findings give a picture of the scholars' collocational treasure. Apparently, both groups of the authors used a limited range of lexical collocation which were probably those collocations they were most often exposed to. This signals the fact that although most of the academic writers are familiar with the notion of collocation, they cannot use collocations creatively and native-like. They use only those collocations that they are familiar with. Kjellmer (1991) claimed the fact by comparing native and non-native speakers and asserted that:

> In building his utterances, he [the native speaker] makes use of large prefabricated sections. The learner, on the other hand, having automated few collocations, continually has to create structures that he can only hope will be acceptable to native speakers [. . .]. His building material is individual bricks rather than prefabricated sections.

> (Kjellmer 1991: 124)

Moreover, sometimes this limited range of used collocations could be the result of L1 effects of the authors, which was totally ignored in this study due to delimited scope of the study. L1 effect could appear in different forms, one of which is word-for-word transfer of the collocations from L1 to L2. In this case, the authors may be more inclined to use those collocations which have equivalences in their L1. As the results depicted, some types of lexical collocations were used more than the others. One assumption could be that these collocations may have an equivalent in the author's L1. To see whether this was the L1 effect and if it was, how it affected the use of lexical collocations, each single collocation should be examined to see whether it has a word-for-word equivalent in the author's L1.

The concluding point of this section that is related to the previous notion mentioned in the above paragraph, concerns the errors that L1 may cause when producing collocations in L2. Since non-natives cannot learn a wide range of prefabricated collocations to use them authentically, they may rely on their L1 in producing collocations, as Kjellmer (1991: 124) pointed "his [the learner's] building material is individual bricks rather than prefabricated sections" which could be the transfer of word-for-word equivalence of L1 collocations. This may lead the authors to produce collocations that are not correct in L2. Since errors were not analyzed in this study, no definite assertion can be made on the role of L1 in production of incorrect collocations in L2. Though, this could be taken into account as an assumption for further research that could be analyzed in future studies.

6.1. Implications of the Study

The findings of this study have both theoretical and pedagogical implications for foreign or second language learning, particularly for foreign/second language writing. At the theoretical level, analyzing Iranian research articles and identifying the most frequent collocations and their types in comparison to those used in Norwegian research articles can lead to developing a specific profile of some writing problems for both Iranian and Norwegian researchers to avoid overuse or underuse of these word combinations.

Concerning the pedagogical implications, given the representation of the corpus, it would be appropriate to expect the findings of this research to be generalized in different areas of language teaching, especially in teaching writing to graduate/higher education students as well as scientific researchers in different areas. The study expands the application of the results to research articles, thus adding to the ever-evolving knowledge of how writing in disciplines can be understood as having predictable use of words. More specifically, the results of this research will be useful to university instructors who want to help students/researchers in achieving a satisfactory level in writing research articles in English.

At the practical level, understanding and awareness of the criteria and key issues relevant to collocations in academic texts provides insight into standards and requirements that scholars, particularly non-native English writers, must meet to improve the quality of their papers and improve the chances of their articles being published in prestigious international journals. The results of this study can be useful in designing tasks and materials for teaching English language writing which focus not only on grammar but also on the skills in writing research articles. The findings can also help to enrich the materials presented in research courses at various levels of higher education to bachelor, master or even PhD degree students.

6.2. Limitations of the Study and Directions for Further Research

No study is perfect; all research studies have their own limitations. Although the current study was an attempt to provide an almost inclusive image of the frequent collocations used by Iranian scholars in comparison to their Norwegian counterparts in research articles, there are limitations to the study which should be acknowledged.

First, the results of the study are only based on frequency-based analysis. It is suggested that other experimental approaches which benefit from human subjects, be employed in order to broaden the scope of the research in this area.

Second, using larger corpora than the corpora used in this study may reveal other results which were not detected in the present study, because one of the main limitations of this study was finding articles published in internationally recognized journals and confirming the authors' mother tongue. Hence, it is highly suggested that similar studies be carried out by making use of larger corpora and different genres.

Third, since the present study compared the use of collocations in research articles written by two non-native groups of scholars, it is suggested that the results of this study be compared to the results of other studies which incorporate any corpora of native English writers to check whether there are any similarities or differences in terms of usage of collocations by native and non-native users. It should be noted that being capable of producing a large amount of collocations does not necessarily imply native-like fluency and underlying intuition of using collocations (Siyanova and Schmitt 2008: 429-458).

Fourth, this study was delimited to identification of only two-word collocations. In order to obtain more comprehensive results, further research using other software capable of detecting and analyzing word combinations with a higher span range as well as grammatical collocations is suggested.

Fifth, as it was addressed in the first chapter of this study, one of the limiting factors in analysis of the results was the statements that were quoted by the authors and did not originally belong to them. Though it was impossible to find all of them, particularly short and direct quotes, it was tried to delete as much quotations as possible from both the corpora. Therefore, in future studies one could examine the impact of quotations in the corpora texts on the final list of recognized collocations by comparing the results obtained from each corpus with and without quotations.

In addition, analyzing errors was not in the scope of this study. Non-natives mistakenly use many combinations as collocations while actually they are not true collocations in their L2. Therefore, it could be an interesting topic for further research to detect and analyze the number of errors and compare them with the number of true collocations.

Moreover, in this research it was decided to include articles published in the *Journal of Pragmatics* due to having consistent editorial policies and style preferences. Other research can be conducted to explore collocations using a different and wide range of sources to compile the corpora. This might reveal some interesting results in comparison to the findings of the present study.

Finally, regarding the contrastive study conducted between the two corpora, it must be noted that the attempts made to explain similarities and differences may be influenced by the interdisciplinary articles' topics. Consequently, future research is obviously required to corroborate the findings of the present study.

References

Printed Sources

Ackermann, K., & Chen, Y. H. (2013). 'Developing the academic collocation list (ACL) – A corpus-driven and expert-judged approach.' *Journal of English for Academic Purposes,* 12, 235–247.

Ahangari, S., & Zununi Vahed, F. (2013). 'Collocation types in Interchange series and high school books.' *Journal of Basic and Applied Scientific Research,* 3(2), 367-374.

Almela, M., Cantos, P., & Sánchez, A. (2013). 'Collocation, co-collocation, constellation... any advances in distributional semantics?' *Procedia-Social and Behavioral Sciences*, *95*, 231-240.

Alsakran, R. A. (2011). The productive and receptive knowledge of collocations by advanced Arabic-speaking ESL/EFL learners. Doctoral dissertation, Colorado State University.

Aisenstadt, E. (1979). 'Collocability Restrictions in Dictionaries.' *ITL International Journal of Applied Linguistics*, (45), 71-74.

Bahardoust, M. (2012). 'Lexical collocations in writing production of EFL learners: A study of L2 collocation learning.' *The Iranian EFL Journal*, *8*(2), 185-200.

Benson, M., Benson, E. & Ilson, R. (1986). *Lexicographic Description of English.* Netherlands: John Benjamins Publishing Company.

Benson, M., Benson, E., & Ilson, R. (1997). *The BBI Combinatory Dictionary of English Word Combinations* (Rev. edition). Amsterdam: John Benjamins Publishing Company.

Bhumadhana, S. (2010). A study of the use of academic verb collocations and English writing ability of undergraduate English-major students. Doctoral dissertation, Chulalongkorn University.

Bradley, H. (1922). *The Making of English*. Macmillan and Company, Limited.

Brillinger, D. R. (2004). 'Some data analyses using mutual information.' *Brazilian Journal of Probability and Statistics*, 163-182.

Brown, D. F. (1974). 'Advanced vocabulary teaching: The problem of collocation.' *RELC Journal*, *5*(2), 1-11.

Carter, R. (1998). *Vocabulary: Applied Linguistics Perspectives*. London: Routledge.

Chen, Y., & Baker, P. (2010). 'Lexical bundles in L1 and L2 academic writing.' *Language Learning & Technology,* 14(2), 30-49.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cortes, V. (2004). 'Lexical bundles in published and student disciplinary writing: Examples from history and biology.' *English for Specific Purposes*, 23(4), 397-423.

Cowie, A. P. (1986). 'Collocational Dictionaries – A comparative view.' In Murphy (ed) *Proceedings of the Fourth Joint Anglo-Soviet Seminar on English Studies.* London: The British Council, 61-69.

Cowie, A. P. (1994). 'Phraseology.' In R. E. Asher & J. Simpson (eds.), *The Encyclopedia of Language and Linguistics*, Oxford: Pergamon Press. 3168-3171.

Cowie, A. P. (ed). (1998). *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.

Darabi, A. (2012). 'On the possible relationship between receptive and productive knowledge of collocational patterns among Iranian TEFL university students.' *The Iranian EFL Journal*, *7*(2), 114-127.

Darvishi, S. (2011). 'The investigation of collocational errors in university students' writing majoring in English.' *IPEDR, 18*, 52-56.

Dash, N. S. (2008). *Corpus Linguistics: A General Introduction*. Pearson Education India. 1-22.

Decarrico, J. S. (2001). 'Vocabulary learning and teaching.' *Teaching English as a Second or Foreign Language*, 3, 285-300.

de Souza Hodne, L. C. (2009). Collocations and teaching: Investigating word combinations in two English textbooks for Norwegian upper secondary school students. Master's thesis, The University of Bergen.

Deuter, M. (ed). (2008). *Oxford Collocations Dictionary: for Students of English*. Oxford: Oxford University Press.

Dunning, T. (1993). 'Accurate methods for the statistics of surprise and coincidence.' *Computational Linguistics*, *19*(1), 61–74.

Eftekhari, M., & Rahimi, A. (2011). 'The effect of delexicalization of common verbs on the collocational competence of Iranian EFL students.' *Procedia Social and Behavioral Sciences, 15,* 3941–3946.

El-Dakhs, D. A. S. (2015). 'The lexical collocational competence of Arab undergraduate EFL learners.' *International Journal of English Linguistics*, *5*(5), 60-74.

Ellis, N. C. (2008). 'The periphery and the heart of language.' In S. Granger & F. Meunier (eds.), *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins Publishing Company. 1-13.

Fan, M. (2009). 'An exploratory study of collocational use by ESL students – A task based approach.' *System, 37*, 110-123.

Farrokh, P., & Mahmoodzadeh, K. (2012). 'Iranian learners' knowledge of English grammatical collocations of gerund.' *Journal of Basic and Applied Scientific Research, 2*(11), 11834-11844.

Firth, J. R. (1957). 'A synopsis of linguistic theory, 1930-1955', *Studies in Linguistic Analysis: Special Volume of the Philological Society*. Repr. In E. Palmer (ed) 1968, *Selected Papers of J. R. Firth*. Oxford: Basil Blackwell, 1-32.

Firth, J. R. (1957). *Papers in Linguistics 1934–1951*. London: Oxford University Press.

Fontenelle, T. (1994). 'What on earth are collocations?.' *English Today, 10*(4), 42-48.

Giacomini, L. (2015). 'Context-dependent variation of LSP collocations: A corpus-based analysis.' *Procedia-Social and Behavioral Sciences, 198*(2015), 140-148.

Gitsaki, C. (1996). The Development of ESL Collocational Knowledge. PhD thesis, The University of Queensland, Brisbane, Australia.

Gledhill, C. (2000). 'The discourse function of collocation in research article introductions.' *English for Specific Purposes 19(2)*, 115-135.

González, A. O., & Ramos, A. M. (2013). 'A comparative study of collocations in a native corpus and a learner corpus of Spanish.' *Procedia-Social and Behavioral Sciences, 95*, 563-570.

Greenbaum, S. (1974). 'Some verb-intensifier collocations in American and British English.' *American Speech*, *49*(1/2), 79-89.

Gulec, N., Gulec, B.A. (2015). 'Lexical collocations (verb + noun) across written academic genres in English.' *Procedia-Social and Behavioral Sciences, 182*(2015), 433-440.

Halliday, M.A., McIntosh, A., & Strevens, P. (1964). *The Linguistic Sciences and Language Teaching*. London: Longman.

Halliday, M. A. (1966). 'Lexis as a linguistic level.' In C. E. Bazell, J. C. Catford, M. A. Halliday, & R. H. Robins (eds.), *In memory of JR Firth*. London: Longman. 148-162.

Hassanabadi, S. (2003). 'A study of the learning of English lexical and grammatical collocations by Iranian EFL learners.' *Journal of Faculty of Letters and Humanities (Tabriz),* 46(187), 45-59.

Hidalgo, E., Quereda, L., & Santana, J. (eds.) (2007). *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi.

Hill, J. (2000). 'Revising priorities: From grammatical failure to collocational success.' *Teaching Collocation*, 47-69.

Hong, A. L., Rahim, H. A., Hua, T. K., & Salehuddin, K. (2011). 'Collocations in Malaysian English learners' writing: A corpus-based error analysis.' *3L: The Southeast Asian Journal of English Language Studies*, *17*, 31-44.

Hsu, J-Y. (2002). Development in collocational proficiency in a workshop on English for general business purposes for Taiwanese college students. Unpublished doctoral dissertation, Indiana University of Pennsylvania.

Hsu, J.-Y. (2007). 'Lexical collocations and their impact on the online writing of Taiwanese college English majors and non-English majors.' *Online Submission*. Accessed 10 Jan. 2018. <https://files.eric.ed.gov/fulltext/ED496121>. 1-23.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.

Hyland, K. (2008). 'Genre and academic writing in the disciplines.' *Language Teaching*, *41*(4), 543-562.

Im, Y., 2006. 'The conception and range of collocation.' *International Association of Language & Literature, 36*, 148–181.

Jian, J. Y., Chang, Y. C., & Chang, J. S. (2004). TANGO: Bilingual collocational concordancer. In *Proceedings of the ACL 2004 on interactive poster and demonstration sessions*. Association for Computational Linguistics. (19). Accessed 15 Jan. 2018. <https://dl.acm.org/citation.cfm?id=1219063>.

Kjellmer, G. (1984). 'Some thoughts on collocational distinctiveness.' In J. Aarts & W. Meijs (eds.), *Corpus Linguistics Recent Developments in the Use of Computer Corpora in English Language Research.* Amsterdam: Rodopi.

Kjellmer, G. (1987). 'Aspects of English collocations.' In W. Meijs, (ed), *Corpus Linguistics and Beyond: Proceedings of the Seventh International Conference on English Language Research on Computerized Corpora.* Amsterdam: Rodopi. 133-140.

Kjellmer, G. (1991). 'A mint of phrases.' In Aijmer, K. & Altenberg, B. (eds.), *English corpus linguistics: Studies in honour of Jan Svartvik*. London: Longman.

Kjellmer, G. (1994). *A Dictionary of English Collocations: Based on the Brown Corpus.* Oxford: Clarendon Press.

Laufer, B., & Waldman, T. (2011). 'Verb-noun collocations in second language writing: A corpus analysis of learners' English.' *Language Learning*, *61*(2), 647-672.

Leech, G. (2001). 'The role of frequency in ELT: New corpus evidence brings a re-appraisal.' *Foreign Language Teaching and Research, 33*(5), Beijing: Foreign Language Teaching and Research Press. 328-339.

Lehrer, A. (1974). *Semantic Fields and Lexical Structure.* Amsterdam: North-Holland Publishing Company.

León, M.A.L. (2015). 'Corpus-based contrastive analysis of keywords and collocations across sister specialized subcorpora in the maritime transport field.' *Procedia-Social and Behavioral Sciences, 198*(2015), 526-534.

Lewis, M. (1993). *The Lexical Approach: The State of ELT and A Way Forward*. Hove England: Language Teaching Publications.

Lewis, M. (2000). 'There is nothing as practical as a good theory.' In M. Lewis (ed), *Teaching collocation*, Hove, England: Language Teaching Publications. 10-27.

Lewis, M. (2002). *Implementing the Lexical Approach: Putting Theory into Practice*. Hove, England: Language Teaching Publication.

Lindquist, H. (2009). *Corpus Linguistics and The Description of English*. Edinburgh: Edinburgh University Press. [1-24, 71-90]

Lyons, J. (1977). *Semantics (2 vols.)*. Cambridge: Cambridge University Press.

Manning, C. D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. Cambridge: MIT press.

Marton, W. (1977). 'Foreign vocabulary learning as problem no. 1 of language teaching at the advanced level'. *Interlanguage Studies Bulletin*, 33-57.

Martyńska, M. (2004). 'Do English language learners know collocations?.' *Investigationes Linguisticae*, *11*, 1-12.

McKeown, K. R., & Radev, D. R. (2000). 'Collocations.' *Handbook of Natural Language Processing. Marcel Dekker*, 1-23.

Mel'čuk, I. (1998). 'Collocations and lexical functions.' In A. P. Cowie (ed), *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press. 23-53.

Men, H. (2017). *Vocabulary Increase and Collocation Learning: A Corpus-Based Cross-sectional Study of Chinese Learners of English*. Singapore: Springer.

Mitchell, T. F. (1971). 'Linguistic 'goings-on': Collocations and other lexical matters arising on the syntagmatic record.' *Archivum Linguisticum*, *2*(1), 35-69.

Moehkardi, R. R. D. (2002). 'Grammatical and lexical English collocations: Some possible problems to Indonesian learners of English.' *Humaniora*, *14*(1), 53-62.

Moore, R. C. (2004). 'On log-likelihood-ratios and the significance of rare events.' In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

Munat, J. (ed). (2007). *Lexical Creativity, Texts and Contexts* (Vol. 58). Amsterdam: John Benjamins Publishing Company.

Nakata, T. (2006). 'English collocation learning through meaning-focused and form-focused activities: Interactions of activity types and L1-L2 congruence.' *Proceedings of the 11th Conference of Pan-Pacific Association of Applied Linguistics*, 154-168.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Nattinger, J. R. (1988). 'Some current trends in vocabulary teaching.' In R. Carter & M. McCarthy (eds.), *Vocabulary and Language Teaching*. New York: Longman. 62-82.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.

Nejadansari, D., & Alijanian, E. (2012). 'The effect of planning on the Iranian EFL learners performance of grammatical collocations in speech.' *World Journal of English Language*, *2*(2), 2-9.

Nesselhauf, N. (2003). 'The use of collocations by advanced learners of English and some implications for teaching.' *Applied linguistics*, *24*(2), 223-242.

Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins Publishing Company.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

Park, J. S., Seraku, T., Kiaer, J. (2016). 'Issues in defining/extracting collocations in Japanese and Korean: Empirical implications for building a collocation database.' *Heliyon 2,* (2016), e00189, 1-25.

Pawley, A., & Syder, F. H. (1983). 'Two puzzles for linguistic theory: Nativelike selection and nativelike fluency.' In J. C. Richards & R. W. Schmidt (eds.), *Language and Communication.* London: Longman. 191-226.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Quping, H. (2012). The effects of collocation instruction on non-English major EFL students' use of collocations. Master's thesis, The University of Bergen.

Rassaei, E., & Karbor, T. (2013). 'The effects of three types of attention drawing techniques on the acquisition of English collocations.' *International Journal of Research Studies in Language Learning, 2*(2), 15-28.

Robins, R. H. (1967). *A Short History of Linguistics.* London: Longman.

Shamsudin, S., Sadoughvanini, S., & Hanafi Zaid, Y. (2012). 'Iranian EFL learners' collocational errors in speaking skill.' *Procedia-Social and Behavioral Sciences 70*, 1295-1302.

Sharifi, A., & Jafarpour, A. A. (2012). 'The effect of error correction feedback on the collocation competence of Iranian EFL learners.' *Teaching English with Technology*, *12*(3)*, 3-17.

Shokouhi, H., & Mirsalari, G. A. (2010). 'Collocational knowledge versus general linguistic knowledge among Iranian EFL learners.' *TESL-EJ*, *13*(4), 1-24.

Sinclair, J. M. (1966). 'Beginning the study of lexis.' In C. E. Bazell, J. C. Catford, M. A. Halliday, & R. H. Robins (eds.), *In memory of JR Firth*. London: Longman. 148-172.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Siyanova, A., & Schmitt, N. (2008). 'L2 learner production and processing of collocation: A multi-study perspective.' *The Canadian Modern Language Review, 64*(3), 429-458.

Skehan, P. (1998). *A Cognitive Approach to Language Learning*. Oxford: Oxford University Press.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell publishers.

Stubbs, M. (2002). 'Two quantitative methods of studying phraseology in English.' *International Journal of Corpus Linguistics*, *7*(2), 215-244.

Tajalli, G. (1994). 'Translatability of English and Persian collocations.' In *Proceedings of the second conference on translation*. Tabriz University, Tabriz.

Tekingül, B. (2013). 'Collocation teaching effect on reading comprehension in advanced EFL setting.' *Procedia-Social and Behavioral Sciences*, *70*, 1078-1089.

Thomson, A. J., Martinet, A. V., & Draycott, E. (1986). *A Practical English Grammar.* Hong Kong: Oxford University Press.

Wang, Y., & Shaw, P. (2008). 'Transfer and universality: Collocation use in advanced Chinese and Swedish learner English.' *ICAME Journal*, *32*, 201-232.

Ward, J. (2007). 'Collocation and technicality in EAP engineering.' *Journal of English for Academic Purposes, 6*(1), 18-35.

Wray, A. (2000). 'Formulaic sequences in second language teaching: Principle and practice.' *Applied Linguistics*, *21*(4), 463-489.

Yule, C. U. (1964). *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

Electronic Sources

Collocation Extract (Version 3.06), Accessed 20 Mar. 2018, Available from <http://pioneer.chula.ac.th/~awirote/resources/collocation-extract.html>.

'*Oxford Collocations Dictionary*', Oxford Learner's Dictionaries, Accessed 10 Jan. 2018, <https://www.oxfordlearnersdictionaries.com/definition/collocations>.

Statistical Package for the Social Science (SPSS), IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Appendices

Appendix 1

Research Article Author's Confirmation Email Sample

Dear Sir/Madam,

I am currently working on my master's thesis at the University of Stavanger, with the title of "Investigating Frequency and Type of Lexical Collocations in Applied Linguistics Journal Articles Written in English by Iranian and Norwegian Scholars".

For the purposes of my study, I need to work on two corpora of articles with certain specifications. As your article titled "...", published on ScienceDirect website (link ...) is suitable for my study and in case you are willing to participate, you are kindly requested to provide the following information:

1. Nationality:
2. Mother Tongue:
3. Second Language:

I hereby would like to assure the confidentiality of any information provided as well as proper referencing of your article in the appendices of the study.

I really appreciate your help in advance and am looking forward to hearing from you soon.

Thanks and regards,
Abbas Moosavi
Mobil: +47- **** ****
Email: sa.moazammoosavi@stud.uis.no

Appendix 2

List of Articles used in the two Corpora

<u>Iranian Corpus</u>

Abdi, R., Rizi, M. T., & Tavakoli, M. (2010). The cooperative principle in discourse communities and genres: A framework for the use of metadiscourse. *Journal of Pragmatics*, *42*(6), 1669-1679.

Abdollahzadeh, E. (2011). Poring over the findings: Interpersonal authorial engagement in applied linguistics papers. *Journal of Pragmatics*, *43*(1), 288-297.

Aliakbari, M., Faraji, E., & Pourshakibaee, P. (2011). Investigation of the proxemic behavior of Iranian professors and university students: Effects of gender and status. *Journal of Pragmatics*, *43*(5), 1392-1402.

Allami, H., & Naeimi, A. (2011). A cross-linguistic study of refusals: An analysis of pragmatic competence development in Iranian EFL learners. *Journal of Pragmatics*, *43*(1), 385-406.

Izadi, A. (2015). Persian honorifics and im/politeness as social practice. *Journal of Pragmatics*, *85*, 81-91.

Keshavarz, F., & Ghassemzadeh, H. (2008). Life as a stream and the psychology of "moment" in Hafiz'verse: Application of the blending theory. *Journal of Pragmatics*, *40*(10), 1781-1798.

Mirzaei, A., & Eslami, Z. R. (2013). Exploring the variability dynamics of wedding invitation discourse in Iran. *Journal of Pragmatics*, *55*, 103-118.

Parvaresh, V., Tavangar, M., Rasekh, A. E., & Izadi, D. (2012). About his friend, how good she is, and this and that: General extenders in native Persian and non-native English discourse. *Journal of Pragmatics*, *44*(3), 261-279.

Shariati, M., & Chamani, F. (2010). Apology strategies in Persian. *Journal of Pragmatics*, *42*(6), 1689-1699.

Norwegian Corpus

Borthen, K. (2010). On how we interpret plural pronouns. *Journal of Pragmatics*, *42*(7), 1799-1815.

Dahl, T. (2004). Textual metadiscourse in research articles: a marker of national culture or of academic discipline?. *Journal of pragmatics*, *36*(10), 1807-1825.

Fløttum, K. (2010). EU discourse: Polyphony and unclearness. *Journal of Pragmatics*, *42*(4), 990-999.

Greenall, A. K. (2009). Towards a new theory of flouting. *Journal of Pragmatics*, *41*(11), 2295-2311.

Ljosland, R. (2011). English as an Academic Lingua Franca: Language policies and multilingual practices in a Norwegian university. *Journal of Pragmatics*, *43*(4), 991-1004.

Raanes, E., & Berge, S. S. (2017). Sign language interpreters' use of haptic signs in interpreted meetings with deafblind persons. *Journal of Pragmatics*, *107*, 91-104.

Svennevig, J. (2012). On being heard in emergency calls. The development of hostility in a fatal emergency call. *Journal of Pragmatics*, *44*(11), 1393-1412.

Svennevig, J., & Djordjilovic, O. (2015). Accounting for the right to assign a task in meeting interaction. *Journal of Pragmatics*, *78*, 98-111.