# The Use of Adjective+Noun Collocations by English Non-Native and Native Writers: A bigram-based study

Lovrić, Benjamin

Master's thesis / Diplomski rad

2020

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Humanities and Social Sciences / Sveučilište Josipa Jurja Strossmayera u Osijeku, Filozofski fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:142:012103

Rights / Prava: In copyright/Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2025-03-11**

J. J. Strossmayer University of Osijek

Faculty of Humanities and Social Sciences

Study Programme: Double Major MA Study Programme in English Language and Literature and German Language and Literature – Teaching English as a Foreign Language

Benjamin Lovrić

# The Use of Adjective+Noun Collocations by English Non-Native and Native Writers: A bigram-based study

Master's Thesis

Supervisor: Dr. Višnja Pavičić Takač, Full Professor

Osijek, 2020

J. J. Strossmayer University of Osijek

Faculty of Humanities and Social Sciences

Study Programme: Double Major MA Study Programme in English Language and
Literature and German Language and Literature – Teaching English as a Foreign
Language

Benjamin Lovrić

# The Use of Adjective+Noun Collocations by English Non-Native and Native Writers: A bigram-based study

Master's Thesis

Scientific area: humanities

Scientific field: philology

Scientific branch: English studies

Supervisor: Dr. Višnja Pavičić Takač, Full Professor

Osijek, 2020

Sveučilište J.J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Studij: Dvopredmetni sveučilišni preddiplomski studij engleskoga jezika i književnosti i njemačkog jezika i književnosti

Benjamin Lovrić

# Uporaba kolokacija pridjev+imenica u engleskim tekstovima izvornih i neizvornih govornika: istraživanje dvopojavnica

Diplomski rad

Mentor: prof. dr. sc. Višnja Pavičić Takač

Osijek, 2020

Sveučilište J.J. Strossmayera u Osijeku

Filozofski fakultet

Dvopredmetni sveučilišni diplomski studij engleskog jezika i književnosti i

njemačkog jezika i književnosti – nastavnički smjer

Benjamin Lovrić

# Uporaba kolokacija pridjev+imenica u engleskim tekstovima izvornih i neizvornih govornika: istraživanje dvopojavnica

Diplomski rad

Znanstveno područje: humanističke znanosti

Znanstveno polje: filologija

Znanstvena grana: anglistika

Mentor: Prof.dr.sc. Višnja Pavičić Takač

Osijek, 2020

IZJAVA

Izjavljujem s punom materijalnom i moralnom odgovornošću da sam ovaj rad samostalno napravio te da u njemu nema kopiranih ili prepisanih dijelova teksta tuđih radova, a da nisu označeni kao citati s napisanim izvorom odakle su preneseni.

Svojim vlastoručnim potpisom potvrđujem da sam suglasan da Filozofski fakultet Osijek trajno pohrani i javno objavi ovaj moj rad u internetskoj bazi završnih i diplomskih radova knjižnice Filozofskog fakulteta Osijek, knjižnice Sveučilišta Josipa Jurja Strossmayera u Osijeku i Nacionalne i sveučilišne knjižnice u Zagrebu.

U Osijeku, 23.09.2020

Benjamin Lović, 0122218560

_____
ime i prezime studenta, JMBAG

# Contents

Abstract

The correct and fluent use of phraseological units, such as idioms, collocations, or phrasal verbs is what differentiates the native from the non-native speaker. The present study focuses on the use of collocations, more specifically on adjective + noun bigrams by non-native English writers. This study is based on the analysis of two corpora, one consisting of 80 essays written by Croatian undergraduate students of the English language and literature, and the other consisting of 32 essays, written by English native speakers attending an American university. The results reveal that non-native writers tend to use more collocations with a higher t-score, while native writers tend to use more collocations with a higher MI value. When the results of the two corpora are compared, it can be seen that native and non-native writers use collocations of different strength on the collocational scale. Thus, the results of the present study confirm previous related studies.

Sažetak

Točna i tečna upotreba frazeoloških jedinica, poput idioma, kolokacija ili frazalnih glagola, razlikuje izvornog govornika od govornika kojemu taj jezik nije materinji. Ovaj se rad bavi upotrebom kolokacija, točnije dvopojavnic tipa pridjev + imenica koje su upotrijebili neizvorni govornici engleskoga jezika u pisanoj produkciji. Istraživanje se temelji na analizi dvaju korpusa, jednoga koji se sastoji od 80 eseja hrvatskih studenata na preddiplomskom studiju engleskog jezika i književnosti te drugoga koji se sastoji od 32 eseja engleskih izvornih govornika. Rezultati otkrivaju da u pisanoj proizvodnji neizvorni govornici engleskoga jezika obično rabe više kolokacija s višim t-rezultatom, dok se izvorni govornici češće koriste kolokacijama s većom vrijednošću MI. Kada se usporede rezultati dvaju korpusa, može se vidjeti da se neizvorni i izvorni korisnici engleskoga jezika koriste kolokacijama različite snage na kolokacijskoj ljestvici. Dakle, rezultati ove studije potvrđuju rezultate ranijih povezanih istraživanja.

Ključne riječi: frazeologija, kolokacije, analiza korpusa, izvorni govornici, neizvorni govornici

# 1. Introduction

Phraseology is a study of certain set or fixed expressions used in a language, for example idioms, collocations, phrasal verbs or other types of multi-lexical items. Many researchers have emphasized the importance of phraseology in second language (L2) acquisition. (Fillmore, 1979; Kjellmer, 1991; Pawley & Syder, 1983). Bestgen and Granger (2014) claim that the correct use of phraseology is one of the main aspects which can distinguish native from non-native speakers. Non-native speakers, even though they appear to have reached a high level of language proficiency, do not use phraseology as much as native speakers do in natural language.

An important part of phraseology are collocations. Even though there is no universally accepted definition of collocations, they are usually referred to as words which occur together more times than by chance. Collocations consist of two parts: a base and a collocate, which has a direct influence on the base. The present study focuses on collocations, more specifically only on adjective + noun collocations found in the corpus of texts written by non-native students of English.

There have previously been numerous studies (Allen 2009; Groom 2009, Ishikawa 2009) on corpora of learner texts but they have usually focused only on the frequency of n-grams in a corpus of learner texts and then compared it to a native speaker corpus. Durrant and Schmitt (2009) have used a different approach and analysed n-grams on the basis of two different statistical measures: the MI (mutual information) and t-score. The MI highlights collocations made up of high frequency words, while the t-score highlights collocations made up of low frequency words. Durrant and Schmitt (2009) have compared a native corpus to a corpus compiled of non-native learner texts. Their results showed that native writers tend to use more low frequency combinations than non-natives, whilst the non-native learners tend to use more high frequency combinations. The result of their study was also confirmed by Bestgen and Granger (2014) who analysed a corpus of EFL writers and then compared the use of collocations by intermediate non-native writers to that of advanced non-native writers. The present study will expand on the study by Durrant and Schmitt (2009) by analysing adjective + noun collocations. More specifically, the study will focus on comparing bigrams found in the corpus of texts written

by learners of English in Croatia, and those in the corpus of texts written by native users of English.

This paper is divided in two parts. The first part provides the theoretical background and a review of previous related research. Thus, the first section of the first part focuses on phraseology and the importance of phraseology in foreign language learning. The second section discusses definitions of collocations and adresses the importance of teaching collocations. The third section reviews the previous research on the topic of collocations and explains in more detail the studies which the present study is based on. The second part of the paper reports on the present study conducted with the aim to analyse and compare the use of adjective + noun collocations in argumentative essays written by native and non-native writers. This will be followed by the results of the study and by the discussion. The last part of the paper brings the conclusion, implications for teaching and directions for future studies.

## 2. Phraseology

Phraseology is a study of certain set or fixed expressions used in a language, for example idioms, collocations, phrasal verbs or other types of multi-lexical items. Phraseological units are very important when learning a new language because their correct use "is one of the aspects that unmistakably distinguishes native speakers of a language from L2 learners, even those that have attained a high level of proficiency." (Bestgen and Granger, 2014: 2) Native speakers appear to be more fluent because they incorporate already predetermined word combinations in their speech, while L2 learners tend to build sentences one word at a time which leads to the impression that their speech is more fabricated when compared to that of the native speakers. The fact that native speakers incorporate predetermined word combinations is confirmed by Pawley and Syder (1983) who claim that native speakers, due to their constant exposure to the language, have a large stock of memorized sentences and phrases. These are used by native speakers automatically. Because native speakers are in possession of these ready-made sequences, they do not have to compose the sequences word-by-word. Native speakers also have a better understanding of the language and can produce a different or an unexpected variation on the normal usage of the predetermined word combination. This further confirms that native speakers appear to be more fluent when speaking. Numerous other researchers acknowledge that the main factor which can distinguish native speakers and foreign language learners is collocational knowledge (Aston, 1995; Fillmore, 1979; Kjellmer, 1991; Pawley & Syder, 1983). Fan (2009) also confirms that the importance of collocational knowledge in L2 competence cannot be disputed. Non-native speakers who have good collocational knowledge sound and write more native-like.

As has already been mentioned, non-native speakers, even those who achieved a high level of proficiency tend to approach the language one word at a time. Durrant and Schmitt (2009) explain that this approach forces non-native speakers to build structures together word by word in ways that they will only later hope to be correct. If non-native speakers piece structures together word-by-word they usually do not use a high variety of different phrases, or in this case collocations. The non-natives are not in contact with the foreign language as much as the natives and therefore have difficulty approaching the language the same way that natives do. According to Fan (2009) native speakers learn collocations differently than non-native speakers. Native

speakers are exposed to collocations in their speech community, thus learning the collocations subconsciously and gradually as they grow. Non-native learners, on the other hand, are usually exposed to the L2 only in a classroom environment, and therefore do not have the same opportunity as the native speakers. This is further confirmed by Farghal and Obiedat (1995) in their study on L2 learners and student translators, who came to the conclusion that student translators and L2 learners are not able to understand and use collocations correctly simply because they are not being made aware of collocations as an important part of multi-word units. The L2 learners and student translators are exposed to L2 only in the classroom environment where vocabulary is taught only as single lexical items, which in turn leads to lexical incompetence of L2 learners.

## 2.1. Collocations

There is no accepted definition about what exactly collocations are. Evert (2007) explains that the concept of collocations is one of the most controversial notions in linguistics. There is a shared belief that certain words have a tendency to co-occur with other certain words in natural language, yet there is not a definition which is accepted by all researchers. There have been many theories and explanations on how exactly collocations should be defined. J. R. Firth (1957) introduced the term "collocations" for characteristic and frequently recurrent word combinations, arguing that the meaning and usage of a word (the node) can to some extent be characterised by its most typical collocates. (Firth, as cited in Evert, 2007: 1). Evert (2007) also argues that the collocations from the Firthian standpoint are interpreted as empirical statements about the predictability of certain word combinations. The definition by Firth was only accepted and formalised after his death by British linguists, who are often referred to as the Neo-Firthian school. This Neo-Firthian approach was not the only approach to collocations, According to Evert (2007) the term "collocations" came to be used in phraseology for some semi-compositional and some already lexically determined word combinations such as *herd of cattle* (rather than *swarm*, *flock* or *school*). Furthermore, it is very difficult, if not impossible to give a clear-cut definition of collocations in the phraseological sense and also differentiate them from restricted word senses. In addition to these two interpretations of the same term, there is also a third meaning of the word collocation in the field of computational linguistics, which uses it as "a generic term for any lexicalised word combination that has idiosyncratic semantic or syntactic

properties and may therefore require special treatment in a machine-readable dictionary or natural language processing system." (Evert, 2007, 2). This definition can be found in most standard textbooks. Evert's (2007) definition of collocations, which also follows the Firthian tradition states that collocations are only a combination of two or more words which have a tendency to co-occur near each other in natural language, meaning that they occur next to each other more often than by chance. This explanation leads to the possibility that words which occur two or more times within a corpus can be considered collocations. Due to this, researchers must decide the minimum frequency of co-occurrence within a corpus for a combination of words to be considered a collocation, especially when dealing with larger corpora which can consist of millions of words. Bartsch (2004), on the other hand, defines collocations as "lexically and/or pragmatically constrained recurrent co-occurrences of at least two lexical items which are in a direct syntactic relation with each other" (p. 76).

Benson et al. (1986) claim that there are many fixed, identifiable, non-idiomatic constructions and phrases in natural languages. The constructions and phrases, due to their characteristics, are called recurrent combinations or fixed combinations, more specifically collocations. According to Bahns and Eldaw (1993) there is a distinction between collocations, idioms and free combinations. They argue that the main difference between collocations and idioms is that the meaning of collocations reflects the meaning of their constituent parts; the meaning of idioms is different than that of their constituent parts. Collocations are also used frequently, they spring to mind readily and are psychologically salient. This is what differentiates them from free combinations. Mel'čuk (1998) defines collocations simply as a single subclass of set phrases. He goes on to explain that researchers cannot observe purely collocations, but must also in turn compare them to other set phrases that are not collocations. Bazzaz and Samad (2011) argue that collocations are an important part of communicative competence. The correct usage of collocations is what differentiates fluent speakers from those who have a lower fluency. Furthermore, collocations are a group of words which fit together intuitively at syntagmatic and paradigmatic levels. Due to this, they are examined in many areas of linguistics, for example phraseology, semantics, systemic linguistics, morphology, syntax, lexicography, and corpus linguistics. Bazzaz and Samad (2011) also explain the lexical view of collocations as a linear and syntagmatic co-occurrence, arguing that collocations are argued as structural word phrases and are syntactically restricted. On their semantic level, collocations can be observed as language

chunks which have a semantic restriction. An example would be the greeting *good morning* which cannot be substituted by **nice morning*.

Several dictionaries have also made a distinction between different types of collocations. One example is The BBI Dictionary of English Word Combinations by Benson, Benson and Illson (1997) who argue that there are two different types of collocations: grammatical and lexical collocations. A grammatical collocation is a phrase which consists of two parts: a content word (a noun, a verb, or an adjective) and a grammatical word, for example a preposition or a grammatical structure (a clause or an infinitive). Some examples of a grammatical collocation are *account for*, *by accident*, and *argument about*. Lexical collocations exclude prepositions, infinitives or clauses and they contain combinations of nouns, verbs, adjectives, and adverbs. Some examples of lexical collocations are *affect deeply* and *bees buzz*. As can be seen from the examples, there are numerous definitions and explanations of what exactly collocations are. A general explanation would be that collocations are words which co-occur more times than by chance and which consist of two parts: a base and a collocate.

## 2.2. L2 collocational competence

Even though many students have a very good communicative competence, they often lack collocational competence. Collocational competence is the ability to use the right collocation in the right context in order to be understood better. Lewis (2009) explains that the lack of collocational competence is usually what forces non-native language users into grammatical errors. Because they do not know the correct collocation, non-native language users tend to create longer utterances. This in turn results in teachers correcting the grammar mistakes, failing to realise that the grammar mistake has occurred due to the lack of collocational competence. Many researchers and teachers have recognized that collocations must be taught in the classroom. One of the first advocates of the implementation of collocations in second language acquisition was Hornby (1974), who even included collocations in the dictionary for advanced learners of English. Brown (1974) also recommends that collocations should be introduced in teaching English as a foreign language. He claimed that this would result in students knowing more collocations, which in turn would make learners more fluent when speaking; learners would be able to read faster and would be able to express themselves better when writing. Bestgen and Granger (2014) even argue that a change of teaching priority must be

made in order for the learners to attain a high level of phraseological competence. Rao (2018) claims that collocations should be an indispensable part of vocabulary teaching. The collocations provide a new and an interesting way for the students to memorize new words. It is necessary for the learners of English to recognize that collocations are an integral part in word acquisition, because the context and the association of collocations provide connections which enable learners to better understand the meaning of a word, and to easily include it in their vocabulary. Therefore, teachers of English must attach sufficient importance to the teaching of collocations, especially when learners are learning new vocabulary. (Rao, 2018)

In their study of German advanced EFL learners, Bahns and Eldaw (1993) came to the conclusion that the knowledge of collocations is necessary for full communicative mastery of English and that collocations should be taught. The subjects of their study used an unacceptable collocate twice as many times as they used an unacceptable general lexical word. Bahns and Eldaw (1993) attributed this to the fact that collocations had been neglected in second language acquisition. Their final conclusion was that not all collocations should be taught. Only those collocations which cannot be paraphrased should be directly taught, whilst the collocations which can be circumvented should not be a priority for teachers. Bazzaz and Samad (2011) state that collocational knowledge can be more beneficial to students than the knowledge of individual words because the knowledge of collocations can contribute to the learners' comprehension and production. They conclude that EFL/ESL teachers' task is to raise learners' awareness of lexical collocations and that teachers should use collocations as teaching resources. (cf. Farrokh, 2012)

Knowledge of L2 collocations is directly linked to the knowledge of vocabulary because the larger vocabulary knowledge implies that the learners will be able to use their large vocabulary knowledge to build a bigger variety of collocations. In other words, in order to achieve a high competence in L2, learners need to have a good knowledge of collocations but must also develop a large vocabulary (Fan, 2009). Vocabulary should be taught in collocational contexts. (cf. Fan, 2009; Rao, 2018) Additionally, Fan (2009) argues that teachers should be encouraged to not only teach collocations which have been put forward by different schools of linguist but they should also be teaching different collocations which they think are relevant to the teaching context and which would benefit both the knowledge of collocations and the general vocabulary knowledge. Farghal and Obiedat (1995) claim that collocations are the most needed and useful genre of

prefabricated speech and that they should be taught explicitly as other multi-word units. Fan (2009) argues that teachers should insist on the usage of collocations and should find a way for the students to use the newly introduced collocational knowledge in the classroom. Only through active usage and multiple encounters with the collocations will learners be able to develop the receptive and productive knowledge of collocations. This in turn will encourage the learners to use L2 collocations in their natural speech or writing. Similarly, Farrokh (2012) argues that "a language can neither be adequately understood nor fluently produced on a word-by-word or purely grammar-focused basis" (p. 70) and that the teachers should encourage the students to use collocations in order to communicate more efficiently.

## 2.3. Teaching collocations

As it has been previously stated, lack of collocational knowledge needs to be directly addressed. This chapter will provide some examples of how teachers can implement collocations in their classroom. First of all, Lewis (2008) warns that teachers have to decide themselves what is important and what is not important and should prepare supplementary materials which will help the learners. Lewis (2008) argues that repetition is a very important variable when teaching a new word or a collocation. He claims that a new lexical item should be repeated a minimum of seven times in different situations in order for the learner to fully understand it and add it to his active vocabulary. One of the recommended methods for teaching collocations is the contrastive analysis. The main strategy of this approach is making a list of collocations in the learners L1 and creating an equivalent list of collocations in the target language. Another method which can be used for teaching collocations, suggested by Lewis (2008), is chunking. This method enables the learner to recognize parts of texts as chunks, and to learn these parts of texts as chunks. When recalling a word, the learner will recall the whole chunk, not only the single word, thus using the new word in a correct manner, usually with another word, thus using collocations. Lewis (2009) recommends that teachers use the opportunity to represent a word as a part of different collocations (take an exam, pass an exam, fail an exam, re-take an exam, scrape through an exam). By using this approach the teacher enables the learners not only to be able to name the word *exam* but also to talk about it in different situations by using collocations. Another example of an exercise suggested by Lewis (2009) would be to ask the students for synonyms of the adjective when they use an adjective + noun collocation. By doing this the teacher presents

different adjectives which can collocate with the original noun. Another example by Lewis (2009) is to present the students with different collocations and to instruct them to use these collocations in a writing task. This encourages them to use the collocation correctly and subconsciously add this collocation to their active vocabulary. Lewis (2008) recommends different types of exercises which the teachers can use to teach multi-word units. These are: identifying chunks, matching, completing, categorising, sequencing, and deleting. These exercises encourage the learner to repeat the words in a new context which in turn makes sure that the learner becomes aware of different situation in which the word can appear. If the focus is strictly on collocations, the teachers should make sure that they use words which appear as parts of collocations and to use those collocations.

Lewis (2008) does not claim that other types of activities cannot be used to teach multi-word units. In order for the other types of activities to be beneficial to collocational knowledge, they need to be adapted to give them a collocational focus. The examples listed above are only a fraction of possibilities the teachers have at their disposal. Every activity and task can be adapted in such a way that it benefits the collocational knowledge of the students.

## 2.4. Previous relevant studies

The field of phraseology has always been an interesting topic for many researchers. Their studies have usually focused on identifying phraseological units and how students use phraseological units. Jiang and Yamashita (2010) analysed the influence of L1 on the acquisition of L2 collocations. Their subjects were Japanese ESL users and EFL learners acquiring English collocations. The results showed that L1 and input have a significant influence on the knowledge and usage of L2 collocations. Bazzaz and Samad (2011) conducted a study on Iranian EFL learners. The focus of the study was on the use of verb + noun collocations in writing stories. They compared the general knowledge of collocations to the use of verb + noun collocations. The result of the study revealed a strong connection between the collocational knowledge and the use of collocations in the writing tasks.

The technological advancement has allowed researchers to look at phraseology from a quantitative approach. Different computer programs allow researchers to analyse larger groups of text, and to analyse only a certain part of the works in which the researchers are interested. One of the most popular techniques used is the extraction of n-grams which are the recurring

sequences of words within a text or texts. A lot of researchers (Allen 2009; Groom 2009, Ishikawa 2009) have analysed n-grams in L2 learner corpora and then compared it to those of native speakers. Bestgen and Granger (2014) argue that the majority of studies which use this method focus almost exclusively on the overall frequency of n-grams in a corpus of learner writing and/or speech. The results are then compared to the results of a comparable native corpus, which in turn uncovers different patterns of mostly overuse and underuse, in some cases also misuse of n-grams. Lukač and Pavičić Takač (2013) analysed a corpus consisting of essays written by higher intermediate English learners and then compared the usage of statistically significant adjective-noun collocations to the usage of the same collocations in a native corpus. One of their conclusions was that non-native speakers tend to overuse general adjectives. Furthermore, they concluded that non-native speakers use a smaller number of adjective-noun collocations when compared to native speakers.

Durrant and Schmitt (2009) used a completely different approach to analysing a corpus and the n-grams within it. They recognize that the learner corpus is not one long text; it is a large group of individual texts, which can be analysed differently. In addition to raw frequency of the word combinations, they analyse the word combinations on the basis of their association scores in a large native reference corpus using two different statistical measures: MI (mutual information) and t-score. These two measures are described as follows:

> rankings based on t-scores tend to highlight very frequent collocations (and so are very similar to rankings based on raw frequency), MI tends to give prominence to word pairs which may be less common, but whose component words are not often found apart. (Durrant and Schmitt, 2009: 167).

What this means is that collocations which have a high t-score appear often in text but are made of words which also occur very often in the text, i.e. have a high frequency. The collocations which have a high MI value do not appear a lot of times in the text, but when they do appear, they usually appear as a part of a collocation. Bestgen and Granger (2014) argue that both of these two statistical measures should be used, because the use of two different statistical measures, in this case the MI and t-score, makes it possible to highlight two opposite aspects of phraseology. The corpus with high MI values and low t-score will mean that learners tend to use low frequency words, while the corpus with low MI values and high t-score will mean that

learners tend to use high frequency words when building word combinations or collocations. This kind of analysis is important for understanding the development of L2 collocational competence. If only one statistical measure was examined, the results would be inconclusive because they would not show the complete difference between the native and non-native corpora. The comparison of two different statistical measures allows the researches to better define the different aspects of L2 acquisition and why the exposure to collocations is important for a better understanding of the language and a better fluency.

Durrant and Schmitt (2009) compared the usage of collocations by native and non-native writers and came to the following findings: their first finding was that native writers tend to use more low-frequency combinations than non-natives. This result was found consistently across the whole corpus. The second finding was that non-native and native writers use approximately the same amount of collocations with very high t-scores. Interestingly, they also concluded that non-native learners tend to repeat certain collocations which they know by heart. This showed that non-native learners tend to overuse collocations with high t-scores in comparison to native learners. The third finding was that non-native writers tend to significantly underuse collocations with high MI values in comparison with native learners. This further confirms the theory that collocations are one of the main distinctions between the native and non-native speakers. Bestgen and Granger (2014) also compared the use of collocations in two different corpora, one by intermediate and one by advanced non-native writers. Their overall results confirm the results obtained in the study by Durrant and Schmitt (2009). Bestgen and Granger (2014) confirmed their hypothesis that intermediate learners use low-frequency but strongly associated collocations less frequently than advanced learners, and that intermediate learners use high-frequency collocations more frequently. The main difference between the two afore-mentioned studies is the size of the corpus and the scope of the research. Bestgen and Granger analysed a bigger corpus and focused on two types of collocations, Noun+Noun collocations and Adjective+Noun combinations.

The present study will analyse and compare collocations written by native and non-native writers. The non-native writers are Croatian learners of English as L2. To our knowledge, there has not been a study focusing on n-grams in a Croatian learner corpus. The present study will

thus fill this gap and allow other researchers to compare the results to those of other non-native learners.

# 3. Methodology

## 3.1. Aim and research questions

This study aims to analyse and compare the use of adjective + noun collocations in argumentative essays written by native and non-native writers.

Three research questions have been set which the study will answer. Those are:

RQ1: How strong are the adjective + noun collocations used by non-native writers?

RQ2: How strong are the adjective + noun collocations used by native writers?

RQ3: Is there a difference between the type of adjective + noun collocations, in relation to their MI value and the t-score, by native and non-native writers?

The hypotheses have also been set. Those are:

1. The adjective + noun collocations used by non-native writers are of medium strength on the collocational category of bigrams (cf. Table 1).

2. The adjective + noun collocations used by native writers are of high strength on the collocational category of bigrams (cf. Table 1).

3. The native writers use collocations with higher MI values and the non-native writers use collocations with higher t-score.

## 3.2. Corpus

This study uses two different corpora, one composed of essays written by native writers and the other composed of essays written by non-native writers. Both groups of writers were given the same instructions. The students were instructed to write an argumentative essay consisting of 200 to 250 words, the topic being "Life in the city". The specific instruction was to write the positive and the negative sides of the life in a city. The essays were hand-written by the students, and then typed to a digital form so that they could be analysed using different programs. The essays were not changed, nor have any errors been corrected.

### 3.2.1. Non-native corpus

The corpus of essays written by non-native writers used in this study consists of a total of 80 essays written by first year students of the Faculty of Humanities and Social Sciences in Osijek. All of the students study English and literature as one of their majors. The essays were collected within the *KohPiTekst* project. The *KohPiTekst* project is a project funded by the Croatian Science Foundation and conducted at the Faculty of Humanities and Social Sciences, University of Osijek. The total count of the words in this corpus is 19,576. In order to standardize the results, the number of collocations per 100 words has been calculated. The non-native corpus contains a total of 2,069 collocations per 100 words.

### 3.2.2. Native corpus

The corpus of essays written by native writers used in this study consists of a total of 32 essays written by students of an American university. All of the students were voluntary participants. The total count of the words in this corpus is 7,537. The discrepancy in the subcorpora sizes is due to the lack of essays written by native writers. In order to standardize the results, the number of collocations per 100 words has been calculated. The native corpus contains a total of 2,070 collocations per 100 words.

### 3.3. Procedure

In order to analyse the corpora, the Sketch Engine software was used. Sketch Engine is a text analysis and corpus manager software developed by Lexical Computing Limited. The software allows the user to upload a number of texts in a digital format and to create a corpus consisting of those texts. After a corpus has been created, the software allows the user to analyse the corpus in great detail. The user can extract n-grams, analyse concordance, compare collocations, look for synonyms etc. The task of the present study was to analyse the adjective + noun collocations. In order to achieve this, the first task was to use Sketch Engine and extract all the nouns in the corpus. In order to have a more consistent result only nouns which occur five or more times in the corpus have been selected. The non-native corpus was the first one to be analysed. The software extracted a list of 165 nouns which have an occurrence of five times or more in the corpus. The native corpus was then analysed. The software extracted a list of 77 nouns. The software does not allow the user to examine the singular and plural forms of the same noun as different nouns. Because of this, the plural and the singular form of the noun will be

presented as an individual result. Each noun was then individually analysed. The software allows the user to see the concordance of the noun, which means that the user can see in which sentence the noun is used and what is its position in the sentence. After a concordance of the noun has been presented in a form of a list, the user can search the concordance for collocations. Collocations are phrases which occur more times than by chance, and because of that the minimum occurrence of a collocation has been set to three or more. This study focuses only on bigrams, so the software was set to only show bigrams. The software does not allow the user to group degrees of comparison of adjectives. Because of this, there will be different results for the positive, comparative and superlative degrees of the same adjective. The software then generates a list of collocates with the selected noun. Each collocation is shown with its t-score and MI value, which are both automatically generated by the software. Each individual collocation was then written in a table with its t-score and the MI value. In order to check the collocation strength of each collocation a threshold proposed by Bestgen and Granger (2014:7) was used. The thresholds can be found in table 1.

Table 1

*Thresholds used for the association measures*

| Categories of bigrams | MI | t-score |
| --- | --- | --- |
| Non-collocational | <3 | <2 |
| Collocational: Low | ≥3 and <5 | ≥2 and <6 |
| Collocational: Medium | ≥5 and <7 | ≥6 and <10 |
| Collocational: High | ≥7 | ≥10 |

# 4. Results

## 4.1. Non-native corpus

The corpus contained a total of 165 nouns which occur five or more times. Of those 165 nouns there were a total of 60 adjective + noun collocations which occurred more than three times in the whole corpus. Each collocation has an assigned t-score and an MI value. All of the collocations with their t-score and the MI value are presented in table 2.

Table 2

*Adjective + noun collocations identified in the non-native corpus with their t-score and MI value*

| Adjective + noun collocation | Frequency of the collocation | t-score | MI |
|---|---|---|---|
| big city/cities | 47 | 6.48 | 4.19 |
| bigger city/cities | 12 | 3.18 | 3.59 |
| many cities | 7 | 1.28 | 0.95 |
| large city/cities | 5 | 2.01 | 3.33 |
| better city/cities | 4 | 1.08 | 1.12 |
| many people | 30 | 5.21 | 4.34 |
| young people | 11 | 3.26 | 6 |
| new people | 8 | 2.74 | 4.99 |
| better people | 4 | 1.58 | 2.26 |
| interesting people | 3 | 1.68 | 5.01 |
| great people | 3 | 1.5 | 2.91 |
| good life | 6 | 2.24 | 3.56 |
| better life | 5 | 1.9 | 2.74 |
| social life | 4 | 1.93 | 4.81 |
| peaceful life | 3 | 1.66 | 4.52 |
| exciting life | 3 | 1.65 | 4.4 |
| many advantages | 20 | 4.35 | 5.23 |
| smaller place/places | 8 | 2.8 | 6.55 |

| | | | |
|---|---|---|---|
| crowded place/places | 4 | 1.95 | 5.36 |
| urban place/places | 3 | 1.69 | 5.28 |
| many places | 3 | 1.44 | 2.58 |
| many disadvantages | 7 | 2.49 | 4.04 |
| small village | 7 | 2.58 | 5.37 |
| smaller village | 4 | 1.97 | 5.87 |
| big traffic | 3 | 1.59 | 3.63 |
| fresh air | 13 | 3.6 | 8.76 |
| polluted air | 6 | 2.43 | 6.76 |
| bad side | 11 | 3.3 | 7.47 |
| good side | 13 | 3.58 | 7.04 |
| negative side | 8 | 2.82 | 7.86 |
| other side | 9 | 2.94 | 5.61 |
| positive side | 4 | 1.98 | 6.96 |
| many sides | 3 | 1.57 | 3.39 |
| better opportunity/opportunities | 5 | 2.17 | 5.14 |
| good thing/things | 7 | 2.61 | 6.33 |
| small town/towns | 10 | 3.15 | 7.88 |
| smaller town/towns | 6 | 2.43 | 7.27 |
| free time | 10 | 3.16 | 8.85 |
| own peace | 5 | 2.22 | 7.19 |
| personal opinion | 3 | 1.72 | 7.61 |
| rural area | 6 | 2.44 | 8.66 |
| urban area | 7 | 2.64 | 8.09 |
| crowded area | 3 | 1.71 | 6.53 |
| high school | 3 | 1.73 | 8.27 |
| busy lifestyle | 3 | 1.73 | 8.33 |
| bigger chance | 3 | 1.71 | 6.54 |
| better chance | 3 | 1.7 | 5.65 |
| different shops | 4 | 1.99 | 8.03 |
| better education | 3 | 1.71 | 6.14 |

| | | | |
|---|---|---|---|
| many options | 4 | 1.96 | 5.67 |
| many events | 3 | 1.69 | 5.37 |
| better connections | 4 | 1.98 | 6.87 |
| large number | 5 | 2.23 | 9.09 |
| new friend/friends | 4 | 2 | 9.02 |
| easy access | 3 | 1.73 | 9.85 |
| easier access | 3 | 1.73 | 9.02 |
| public transport | 7 | 2.64 | 10.54 |
| big company/companies | 3 | 1.72 | 6.75 |
| public transport | 4 | 2 | 9.93 |
| important factor | 3 | 1.73 | 8.63 |

As can be seen from table 2, there is a great variation in the use of adjective + noun collocations by the non-native writers in the essays. The most frequently used adjective + noun collocation is *big city/cities,* which occurs 47 times in the whole corpus. This collocation has a t-score of 6.48 and an MI value of 4.19, which puts in the medium range on the threshold shown in table 1. The second most frequent collocation is *many people* with its t-score of 5.21 and the MI value of 4.34. Both the t-score and the MI value put it in the low range on the threshold scale. On the other side, one of the collocations with the least occurrences is *easy access* which has a t-score of 1.73, which makes it non-collocational according to the threshold given in table 1, but its MI value of 9.85 puts it high on the collocational range. Another example of such discrepancy is the collocation *public transport* whose t-score of 2.64 puts it in the low range on the threshold scale, but its MI value of 10.54 makes it highly collocational according to the threshold scale. When all the results of the non-native corpus are examined, it can be clearly seen that the collocations which have a high frequency in the corpus also have a bigger t-score but a lower MI value. This is due to the fact that the t-score has a higher value with the collocations which consist of words which occur individually often in the text. For example, the collocation *big city/cities* occurred 47 times in the whole corpus, but the noun *city/cities* occurred a total of 648 times in the whole corpus. Even though the collocation *big city/cities* occurs the most times in the corpus, it does not occur enough many times for it to have a higher t-score. Due to this, there are no collocations with a big enough t-score to place it high on the collocational scale. On the other hand, there are

in total 21 collocations with an MI value of over seven. This means that these 21 have a big enough MI value to be placed high on the collocational scale. This is explained by the fact that these collocations consist of low frequency words in the corpus, and the words which build the collocations usually only occur as a part of the collocation. An example of this is the collocation *public transport*, with the MI value of 10.54, making it the highest MI value in the whole non-native corpus. This is explained by the fact that the collocation *public transport* occurs seven times in the non-native corpus, but the noun *transport* only occurs eight times in the whole corpus. This means that the noun *transport* is only used one time in the whole corpus without the adjective *public*.

## 4.2. Native corpus

The native corpus contained a total of 78 nouns which occurred five or more times. Of those 78 nouns there were 30 adjective + noun collocations which occurred more than three times in the whole corpus. Each collocation has an assigned t-score and an MI value. All of the collocations with their t-score and the MI value are presented in table 3.

Table 3

*Adjective + noun collocations identified in the non-native corpus with their t-score and MI value*

| Adjective + noun collocation | Frequency of the collocation | t-score | MI |
|---|---|---|---|
| big city/cities | 13 | 3.45 | 4.49 |
| large city/cities | 4 | 1.81 | 3.37 |
| different life | 22 | 1.58 | 3.54 |
| some people | 6 | 2.34 | 4.47 |
| many people | 6 | 2.29 | 3.95 |
| new people | 4 | 1.88 | 4.1 |
| more people | 5 | 1.97 | 3.08 |
| most people | 3 | 1.65 | 4.47 |
| professional opportunity/opportunities | 3 | 1.72 | 7.14 |
| great opportunity/opportunities | 3 | 1.71 | 6.29 |
| more opportunities | 4 | 1.87 | 3.95 |

| | | | |
|---|---|---|---|
| many opportunities | 3 | 1.63 | 4.14 |
| many advantages | 5 | 2.17 | 5.14 |
| some advantage/advantages | 3 | 1.67 | 4.92 |
| good place/places | 5 | 2.19 | 5.76 |
| many disadvantages | 3 | 1.69 | 5.4 |
| city living | 7 | 2.47 | 3.93 |
| different culture/cultures | 3 | 1.71 | 6.47 |
| cultural event/events | 3 | 1.73 | 8.05 |
| public transportation | 12 | 1.73 | 8.88 |
| many activities | 4 | 1.97 | 6.27 |
| rural area/areas | 8 | 2.82 | 9 |
| mental health | 3 | 1.73 | 9.46 |
| green space/spaces | 3 | 1.73 | 9.14 |
| human interaction | 3 | 1.72 | 7.82 |
| high rate/rates | 4 | 2 | 8.76 |
| transportation system/systems | 3 | 1.73 | 8.35 |
| small centre/centres | 3 | 1.73 | 8.73 |
| other hand | 4 | 1.99 | 8.2 |
| young adult/adults | 4 | 2 | 9.56 |

Even though the native corpus has almost three times fewer words than the corpus from non-native writers, it has 30 collocations, which is exactly the half from the non-native writers. Because it is a smaller corpus, there is also a smaller variety of collocations which occur in the corpus. Out of the 78 nouns which occur in the corpus, only 21 nouns come as a part of a collocation. The most frequently used adjective + noun collocation is *different life,* which occurs 22 times in the whole corpus. This collocation has a t-score of 1.58, making it non-collocational, and an MI value of 3.54, which puts in the low range on the threshold shown in table 1. Interestingly, this collocation does not occur in the corpus of essays written by non-native writers. The second most frequent collocation with 13 occurrences is *big city/cities* with its t-score of 3.45 and the MI value of 4.49. Both the t-score and the MI value put it in the low range on the threshold scale. This is the collocation which was most frequently used by the non-native

writers. There are 13 collocations which occur only three times in the whole corpus. One of the collocations which occurs only three times in the corpus is *mental health,* which has the t-score of 1.73, making it non-collocational, but its MI value of 9.46 places it high on the threshold shown in table 1. The collocation with the highest MI value is *young adult/adults,* which has an MI value of 9.56, placing it high on the threshold, and a t-score of 2, placing it low on the threshold. A variation of the collocation *public transport*, which has the highest MI value in the non-native corpus, is present in the native corpus. The collocation used by the native writers in this case is *public transportation*, which has a t-score of 1.73 and an MI value of 8.88. What is interesting about this collocation is that it occurs 12 times in the whole corpus, and the noun *transportation* also occurs only 12 times in the whole corpus. This means that the noun *transportation* is used exclusively as a part of the collocation. The MI value of 8.88 which this collocation has can be attributed to the fact that the adjective *public* occurs 19 times in the corpus. Even though this collocation occurs only 12 times in the corpus, it is still the third most frequently used collocation in the whole corpus. Its t-score of 1.73 makes it non-collocational. When all of the results are examined, it can be seen that the native writers, given the size of the corpus, used a higher number of collocations than the non-native writers. As can be seen from table 3, only three collocations occur more than ten times in the whole corpus. Even though there were a total of 30 collocations used in the corpus, only nine of them have a t-score high enough to be considered collocation, the rest are non-collocation according to the threshold in table 1. These nine collocations are also of low strength on the threshold, meaning that the students have not used a significant number of collocations made up of high frequency words. On the other hand, all 30 different collocations have a high enough MI value to be considered collocational on the threshold scale. Twelve collocations are of low strength on the collocational scale, six are of medium strength, and twelve are of high strength. This means that the collocations are made up of low frequency words, and when these words appear in the corpus, they are often a part of the collocation. The most prominent example is the collocation *public transportation*. Another great example is the collocation *young adult/adults* which appears 4 times in the corpus, but the noun *adult/adults* appears a total of 5 times in the corpus. There is only one instance in which the noun *adult/adults* is used without the adjective *young*.

## 5. Discussion

The focus of this study was to analyse and interpret the use of adjective + noun collocations in argumentative essays written by native and non-native writers, and to compare the results.

The results indicate that the first hypothesis can be partially accepted, because the collocations used also are placed in the high range of the threshold seen in table 1. The second hypothesis can also be accepted, because the collocations used are of high strength on the collocational scale, but only if the MI value is observed. Exactly due to this reason, the two different statistical measures had to be used. Finally, the third hypothesis can be accepted, because the non-native learners tend to mostly use high frequency collocations which have higher t-scores, whilst the native learners tend to mostly use low frequency collocations which have higher MI values.

Based on the results shown in this study, it can be concluded that non-native learners use a great variation of adjective + noun collocations when writing an essay, but mostly use collocations with a medium to high T-score. What this means is that the collocations used most frequently in the essays consist of high frequency words. The collocations which do not occur as often in the essays have a higher MI value, which means that they consist of low frequency words. This means that the words which make these collocations are usually found most frequently only as a part of the collocation itself. The present study expands on study done by Durrant and Schmitt, who also compared the usage of collocations by native and non-native writers. The result of their study was that non-native learners tend to overly rely on forms which are common in the language. The fact that they use a lot of collocations with high t-score indicates that they have a preference for very frequent collocations. Durrant and Schmitt (2009) argue that learners tend to quickly learn and use highly frequent collocations, but have issues learning collocations which are less common but strongly associated, for example *densely populated*, *bated breath*, or *preconceived notions*. The results of the present study further confirm this theory.

Native learners also use a high range of different collocations, especially if the size of the native corpus is compared to the size of the non-native corpus. Even though native learners used a lot of different collocations, the collocations were not used very often. On the other hand, the non-native writers used more collocations in their essays. The collocations used by native writers have a very low t-score: out of the 30 collocations, only nine are collocational and all nine are of

low collocational strength. On the other hand, all 30 collocations have a high enough MI value to be considered collocational, and twelve of those are of high strength on the collocational scale. What this means is that, even though native writers do not use as many collocations as non-native writers, the native writers use collocations made of low frequency words. This means that they approach the language differently than non-native writers. The native writers use pre-fabricated speech patterns, confirming the argument by Pawley and Syder (1983) who say that "possession of a large stock of memorized sentences and phrases simplifies his task in the following way. Coming ready-made, the memorized sequences need little encoding work." (p. 208). The fact that all the collocations have a high enough MI value to be considered collocational confirms this argument, because the phrases, or in this case collocations, are used without analysing the individual segments of the collocation. This result is also in line with Durrant and Schmitt's (2009) study, who concluded that native learners use more low-frequency collocations than non-native learners.

The native writers, when writing and deciding which words to write, recall the collocation as a memorised phrase, i.e. the collocation is seen as a single unit. The non-native writers approach the language differently: the collocation is not remembered as a single unit. The non-native writer starts with a word which he wants to use, and then modifies this word, creating in some cases inadvertently a collocation. This also further proves Durrant and Schmitt's theory that non-native writers "piece structures together word-by-word in ways that they can only hope will prove acceptable" (2009: 158). The difference between the approach to the language by native and non-native learners can be explained by the fact that the non-native learners are not exposed to the target language as much as native learners (Fan, 2009). The native learners acquire collocations automatically and subconsciously because they are exposed to them as they grow up, the native learners do not have to learn them by heart. The non-native learners are usually only exposed to the target in a classroom environment and therefore do not have the same chance to learn collocations as the native learners do. The non-native learners are, simply put, not exposed to the target language often enough to learn the collocations in the same way that native learners do which in turn leads to a completely different approach to the language. Durrant and Schmitt (2009) concluded that the advanced phraseology of non-native learners is different from that of the native speakers not because it avoids the formulaic language, but because it underuses the lower-frequency, strongly-associated collocations characterised by high MI values and

overuses high-frequency collocations. The findings in the present study confirm this statement. The non-native writers use more collocations with higher t-score, meaning that they are made of high frequency words, while the native writers use collocations with high MI values, meaning that they are made of low frequency words.

The results of the present study are important because they show that the pattern of usage of collocations by non-native learners is similar in the interlanguage of learners with different L1s. This study was done on Croatian learners of English and is in line with other research which was done on different non-native learners. This proves the theory that learners are not exposed to collocations as often as they should be, which in turn results in their sounding less fluent, even though they have achieved a high level of proficiency.

The limitation of this study was the size of both corpora, especially the native corpus. It would be interesting to conduct another research on larger corpora. This in turn would lead to more reliable results, and the comparison between the native and the non-native corpus would be even more significant. The present study and the studies done by Durrant and Schmitt (2009) and Bestgen and Granger (2014) have only grazed the surface of this part of phraseology. They have created a possibility for a number of follow-up studies. What would be interesting is the comparison between multiple different non-native corpora. An analysis could be done to determine if different non-native learners have a tendency to use different collocations or of different strength on the collocational scale.

# 6. Conclusion

Based on the results presented, it can be concluded that the non-native EFL learners in Croatia most frequently use adjective + noun collocations which are placed in the medium and high range in relation to their respective t-score and MI value. Next, native writers frequently use adjective + noun collocations which are considered collocational according to their MI value. The collocations which have a higher frequency also have a lower t-score, meaning that they are not collocational. Finally, it can be concluded that there is a difference in the usage of adjective + noun collocations by native and non-native learners, when their MI value and the t-score are examined.

The results of the study show the importance of teaching collocations in a classroom environment. The use of phraseology and collocations is very important for fluency of a non-native learner. The difference between the usages of collocations implies that the non-native learners should be exposed to collocations more often and in different situations in order for them to have a higher competence. The present study also confirms the arguments put forward by Brown (1974) and Rao (2018), who agree that collocations should be an integral part of second language acquisition. If a non-native learner aspires to achieve better fluency and to sound more native-like, the non-native learner should focus on learning collocations, which in turn would lead to him appearing more native-like. The teachers should be made aware of the importance of the collocational knowledge and should strive to introduce collocations in the classroom environment. It is clear that the use of collocations is strictly connected to the fluency and more focus should be put on making sure that learners understand the importance of collocations.

# 7. Bibliography

Allen, D. (2009). Lexical bundles in learner writing: an analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education* 1, 105–127. http://park.itc.u-tokyo.ac.jp/eigo/KJEE/001/105-127.pdf (accessed 17 March 2020).

Bahns, J. and Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1), 101-114.

Bartsch, S. (2004). *Structural and Functional Properties of Collocations in English*. Narr: Tübingen.

Bazzaz, F. E. and Samad, A. A. (2011). The use of verb noun collocations in writing stories among Iranian EFL learners *English Language Teaching, 4*(3), 158-164.

Benson, M., Benson, E. and Ilson, R. F. (1986). *The BBI Combinatory Dictionary of English.* Amsterdam: John Benjamins.

Bestgen, Y. and Granger, S. (2014). The use of collocations by intermediate vs. advanced non-native writers: A bigram-based study *International Review of Applied Linguistics in Language Teaching, 52*(3), 229-252.

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

Durrant, P, and Schmitt N. (2009). To what extent do native and non-native writers make use of collocations? *International Journal of Applied Linguistics in Language Teaching 47*(2), 157–177.

Evert, S. (2007). Corpora and collocations. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics. An International Handbook*, article 58 (pp. 1212-1249). Mouton de Gruyter: Berlin.

Fan, M. (2009). An exploratory study of collocatioanal use by ESL students – a task based approach *System 37*(1) 110-123

Farghal, M. and Obiedat, H. (1995). Collocations: a neglected variable in EFL. *IRAL 33*(4), 315-331.

Farrokh, P. (2012). Raising Awareness of Collocation in ESL/EFL Classrooms. *Journal of Studies in Education, 2*(3), 55-74.

Fillmore, J. C. (1979). On fluency. In D. K. C. J. Fillmore and W. Wang (eds.), *Individual differences in language ability and behavior* (pp. 85-101). New York: Academic Press.

Firth, J. R. (1957). A synopsis of linguistic theory 1930–55. In *Studies in linguistic analysis*, 1–32. The Philological Society, Oxford. Reprinted in Palmer (1968), 168-205.

Groom, N. (2009). Effects of second language immersion on second language collocational development. In A. Barfield and H. Gyllstad (eds.), *Researching Collocations in Another Language* (pp. 21–33). Houndmills: Palgrave Macmillan.

Hornby, A.S. (1974*). Oxford Advanced Learner's Dictionary*. Oxford: Oxford University Press

Ishikawa, S. (2009). Phraseology overused and underused by Japanese learners of English: A contrastive interlanguage analysis. In K. Yagi and T. Kanzaki (eds.), *Phraseology, Corpus Linguistics and Lexicography: Papers from Phraseology 2009 in Japan* (pp. 87–100). Nishinomiya: Kwansei Gakuin University Press.

Jiang, N. and Yamashita, J. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations *TESOL Quarterly, 44*(4), 647-668.

Kjellmer, G. (1991). A mint of phrases. In J. Svartvik, K. Aijmer and B. Altenberg, *English corpus linguistics: Studies in honour of Jan Svartvik* (pp. 111-127). New York: Longman.

Lewis, M. (2008). *Implementing the lexical approach: Putting theory into practice*. Andover: Cengage Learning.

Lewis, M. (2009). *Teaching collocation: Further developments in the lexical approach*. Hove: Language Teaching Publications.

Lukač, M., and Pavičić Takač, V. (2013). How word choice matters: An analysis of adjective-noun collocations in a corpus of learner essays. *Jezikoslovlje, 14*(2-3), 385-402.

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

Pawley, A. and Syder F. H. (1983). Two puzzles for linguistic theory: nativelike selection and nativelike fluency. In Jack C. Richards and Richard W. Schmidt (eds.), *Language and Communication* (pp. 191–226). London: Longman.

Rao, V. C. S. (2018). The Importance of Collocations in Teaching of Vocabulary. *Journal for Research Scholars and Professionals of English Language Teaching 7(2)*, 1-8.