VERBAL COLLOCATIONS IN THE SPOKEN ALBANIAN CORPUS IN KOSOVO

Adelajda BAFTIU, Bardh RUGOVA

University of Prishtina "Hasan Prishtina", Kosovo
E-mail: adelajda.baftiu@uni-pr.edu; bardh.rugova@uni-pr.edu

ABSTRACT: This paper aims to examine verbal collocation in the spoken corpus of Albanian in Kosovo. Lexical and semantic components that are repeated in a linguistic context are known as collocations. The verbal collocations occurring inside the parameters of the sentence will be examined in this study. The most common verbs found in the corpus, those used more than 70 times, are used as the basis for this study. Examples include: *vij, bëj, dal, hyj, marr* and *rri*. Except for the verbs *rri* and *hyj* which came out with 16 meanings, the same verbs were presented with more than 30 meanings in the Dictionary of the Albanian Language. Due to their wide range of meanings and ditransitive nature, the verbs *marr* and *bëj* are used in the most consistent collocations. So, they are more likely to be followed by a complement unlike the verbs *rri, hyj, dal* and *vij.*

The work makes a connection between corpus linguistics and semantics. The methods of Baker (2010) and Sinclair (1991) are applied in this investigation. They base their strategy on how frequently the words x and y are displayed next to one another, setting a spacing of five words to the left and five words to the right of the word x. 76 recordings of speakers from various cities in Kosovo make up the corpus of this study. The Partitur Editor application is used to transcribe recordings. The KWIC program is then used to automatically extract collocations as well as vocabulary from the transcriptions.

KEYWORDS: collocations, verbal collocations, semantics, corpus, corpus linguistics, spoken language, spoken corpora

1. Introduction

The vocabulary of a given language consists of a variety of lexical systems, the semantic structure of which can be explained using paradigmatic and syntagmatic semantic relations. Each linguistic unit has its place in a system and derives its function or value from the relationships it makes with other units of the system, according to one of the fundamental principles of structuralism proposed by Saussure (Lyons 2001). This is critical for comprehending the idea of collocations, whose study depends on context because the links between words in a text have a significant impact on how a particular term is understood. For instance, the meaning of the words *marr* (take) and *vesh* (ear) when the two words are used separately and in collocation vary greatly. When these words are collocated, then their meaning is *kuptoj* (understand). Collocation, which literally translates to "joining, grouping," was originally used by Firth in 1957. He believes that speakers use typical lexical choices that are common in colloquial combinations. Firth's theory of meaning makes use of the concept of collocation (cit. According to Gelbukh and Kolesnikova 2013). In a corpus, collocations are thus words that are frequently used in conjunction with one another; one of these words serves as the base, and the other as the collocation. They are divided into grammatical collocations and lexical collocations. Baker (2010, p. 24) explains collocations in relation to corpus linguistics. He gives the definition of collocations as follows *"collocation is a way of demonstrating (relatively) exclusive or frequent relationships between words (or other linguistic phenomena). If two words collocate, then they have a tendency to occur near or next to each other in naturally occurring language use"*. Baker states that there are several different types of collocations, in which words can appear regularly as part of the idiom *(drop off, catch on),* complex noun *(swimming pool),* lexical items *(I don't know).* In these cases, the position of the two collocations in relation to one another is almost always the same, but collocations may have different positions, e.g. *tell* and *story* are collocations because they appear in a range of different grammatical contexts: *tell me a story, story to tell, let the story tell itself, tell a story, that story does not tell us anything*. Collocations are studied from different linguistic perspectives. Each field then has different criteria to distinguish them. Gelbukh and Kolesnikova (2013) define collocations and specify the circumstances in which a word combination meets the criteria for a collocation. Some of them are given in the following list of definitions for collocations (Gelbukh and Kolesnikova, 2013, p. 22): "*Collocations of a given word are statements of the habitual or customary places of that word*" (Firth, 1957), "*Collocation is the co-occurrence of two units in a text within a specified environment"* (Sinclair, 2004), "*Collocation is a combination of two lexical items in which the semantics of one of the lexical items (the base) is autonomous from the combination it appears*

in, and where the other lexical item (the collocate) adds semantic features to the semantics of the base" (Mel'čuk, 1998).

It is clear from the list of definitions provided by Gelbukh and Kolesnikova (2013) that different authors employ different criteria, such as lexical, statistical, functional, and semantic criteria, to identify collocations. The base, which is the word used in its primary and typical sense, and the collocation, which is used to express a meaning different from its typical meaning and conveyed in the combination of free words, are the two components that make up a collocation (Gelbukh and Kolesnikova 2013). The collocation is a word that is semantically dependent on the base word. The cohesion between the collocation elements is strong enough that the base chooses one particular word and not another to express a certain meaning. Gelbukh and Kolesnikova (2013, p. 4) give this definition for collocations: *"Expressions in which one word chooses another one to convey a particular meaning in an unmotivated, unpredicted way are called collocations or restricted lexical co-occurrence".* This study focuses on the interconnectedness between the semantics and corpus linguistics fields. Semantics is referred to as the science that studies meaning, whereas corpus linguistics is the study of linguistic information based on computerized collections of spoken or written texts. Corpus linguistics is useful in this situation to distinguish collocations because it provides the frequency of each word in the corpus and the relationships with the other words in the context. The object of this paper is to examine the verbal collocation in the spoken corpus of Albanian in Kosovo. Verbal collocations are analyzed into their component parts in order to ascertain if they are only used with nouns or with other grammatical elements, as well as to ascertain the meaning that each collocation conveys. The most frequent verbs in the corpus, those with more than 70 occurrences, are used as the basis for this analysis without taking into account the auxiliary verbs. Thus, only lexical collocations are considered. It is determined which word in a collocation bears the primary meaning and which word is semantically dependent once the collocations are gathered and their meanings are examined. For instance, in the phrase *marr vendim* (take a decision), the verb *marr* is semantically dependent on the base word while the noun *vendim* bears the primary meaning. The research question of this paper is: Do these verbs, which are used more often in the spoken corpus, create idioms?

2.   Methodology

There are numerous methods for calculating them due to the multidimensional concept of collocations. This study exclusively considers collocations that occur within sentences and employs statistical criteria to identify them, looking at the four- to five-word space to the right and left of the word x, methods used by Sinclair (1991) and Baker (2010). The calculation approach used in this method, which considers the relative placements of the two words across the corpus, is closely connected to the calculating method which is based on the strength of the relationship between two words. The corpus of this paper includes 76 recordings by speakers of these cities in Kosovo: Ferizaj, Shtime, Shtërpcë Kaçanik, Gjilan, Mitrovica, Opoja and Istog. This corpus used in this study is part of the spoken corpus of Albanian in Kosovo, which is in the transcription phase of the recordings and is not yet published. The corpus is a project of the Academy of Sciences and Arts of Kosovo. All regions of Kosovo are included in the corpus, which seeks to be representative of the language used in Kosovo. The corpus is intended to contain 500,000 words, 2,000 recordings or samples, and each sample plus roughly 250 words. The length of recordings is usually 6 minutes. The interviews were conducted between 2019 and 2021. The corpus solely includes casual spoken texts, such as comments about different events, family conversations, dialogues between speakers who are acquainted with one another and those who are not at all acquainted with one another, and telephone conversations. In order for the speakers to respond uniformly to questions in the many dialects of Albanian spoken in Kosovo, numerous recordings were made through informal interviews. The subjects described personal experiences, showed how they spend a typical working day and talked about their professions. The majority of the interviews took place at participants' homes, or at schools, colleges, stores, squares, cafes, workplaces, autos, etc. Dictaphone, Samsung phone, and iPhone are devices used for recording the interviews.

Following the interview, the subjects answered questions about their place of birth, education level, mother tongue, foreign languages, career, age, and gender on a survey. The transcription of the recordings was done using the Partitur Editor program. Following transcription, each file is saved individually and imported into the KWIC application where it is saved as plain text. As a result, only the transcription of the audio was done using the Partitur Editor application. The KWIC program was also used in conjunction with the analysis because it was more appropriate for this investigation. The KWIC program is then used to extract the vocabulary from the transcripts while automatically extracting the collocations. The primary criteria on which we shall rely are

frequency and the semantic criterion. It was also necessary to look up collocations in the dictionary to determine which words provided a more thorough description of the collocations. This study is based on the premise that verbs having several meanings (semantic nuances) and a higher frequency in the corpus create stable or strong collocations. (chosen from basic words to express a certain meaning), e.g. the verbs *marr (take), bëj (do),* in contrast to the verbs which have a lower frequency and appear with fewer meanings in the dictionary, such as *hyj (enter).*

2.1.   Methods for calculating collocations

The term *multi- word expression* is a general term that refers to phenomena, such as: *cliché, collocation, fixed expression*, *idiom*, *multi-word unit*, *phraseme*, *phrasal lexical item, phrasal lexeme*, *prefabricated chunk, prefab etc*. Hüning and Schlücker (2011) use the term *multi-word expression* as a general term that includes phenomena with different degrees of syntactic fixedness and semantic compositionality. In the list of the phenomena commonly regarded as MWEs they enter: proverbs (*A bird in the hand is worth two in the bush*), quotations (*Shaken, not stirred*) and commonplaces (*One never knows*), metaphorical expressions (*as sure as eggs is eggs),* verbal idioms (*to kick the bucket*), particle / phrasal verbs (*to make up)*, light verb constructions / composite predicates (*to have a look)*, syntactic / quasi noun incorporation (German *Auto waschen* „to wash car), stereotyped comparisons / similes (*as nice as pie*), binomial expressions (*shoulder to shoulder),* complex nominals (*man about town*, *weapons of mass destruction*, *sheep's clothing*, collocations (*strong tea*), fossilized / frozen forms (*all of a sudden)*, routine formulas (*Good morning*). Generally, a distinction is made between fully opaque, non-decomposable MWEs, decomposable MWEs which contain one or several words with an idiomatic meaning, and fully compositional, non-idiomatic MWEs. So, MWEs it is a broader term that also includes collocations.

Using two different principles of interpretation, the principle of open choice and the concept of idiom, Sinclair (1991) demonstrates how meaning develops from the linguistic text. According to the open choice concept, the linguistic text is the outcome of a very large number of complex choices. A large number of options become available at each point where a unit is finished (word, phrase, or clause), and the only restriction is grammaticality. On the other hand, although appearing to be able to be broken down into segments, the idiom principle makes a vast number of semi-ready phrases with single choices available to language users. Collocations, according to Sinclair (1991), illustrate the principle of idiom. Collocation is the occurrence of two or more words within a short space of each other in the text. The usual measure of closeness is the maximum of four words between. Frequently recurring collocations are related to the significance of statistical measures.

Collocations can be calculated in a number of different methods, each of which expresses a distinct type of relation in terms of frequency and exclusivity, as noted by Baker (2010). The simplest way is to count the number of times the word x appears next to the word y, setting the gap as being five words to the left and five words to the right of the word x. Of course, words within the same sentence can be stronger collocations than words that appear in different sentences. The *Mutual Information Score* approach, which Baker also discusses, uses two words to determine how closely related they are in order to determine collocations. This approach takes into account how frequently and seldom the two words appear close to one another across the corpus. Then they get high scores. Their scores will be lower if they frequently appear both together and separately. And even if they only occasionally appear together and typically appear individually, the scores will still be higher low. According to Baker, the third method, *Dice coefficient*, generally reveals the most prevalent lexical collocations: *tells, short, detective, true, success, whole, adventure, read, love*. This way of calculating collocations is based on the type of collocations we want to focus on. For example, we can only be interested in grammatical collocations that have a high frequency, but we can also be interested in pairs of collocations that have a low frequency.

This study uses the statistical method of finding collocations, but always takes into account *Mutual Information Score* and *Dice coefficient*. From what was said above, it can be seen that in language, regardless of the linguistic field, the statistical method is the method that is used the most to find collocations.

3.   Results and analysis

Collocations can be divided into six categories: adjective + noun, noun + noun, verb + noun, adverb + adjective, verb + prepositional phrase, and verb + adverb. This paper deals only with verbal collocations. Gelbukh and Kolesnikova (2013) stress that verbal collocations are those verbal combinations with other words (often with prepositions, nouns, and adverbs), the meaning of which cannot be inferred just from the

meaning of their individual constituent parts. In the corpus of this work of 76 recordings, a word of 7500 words and 57513 word forms are used.

The verbs with the highest frequency in the Albanian spoken corpus in Kosovo are the following:

Verbs with the highest frequency in the Albanian spoken corpus	Use
Vij (ardh) – come	99
Bëj (ba/bo) – do	197
Dal – go out	98
Hyj (hi) – come in	105
Marr – take	76

Table 1. Verbs with the highest frequency in the Albanian spoken corpus

3.1. Collocations of the verb *marr (take)*

The verb *marr* in the selected corpus is used 76 times. The Dictionary's entry for the first definition in Albanian is: "*grab someone or something by hand, etc., grab and hold it: took the book; take the dust with a cloth; / catch, voice: took it by the arm (for hand, drag); they took him hostage.* " 34 different meanings of the verb *marr* are given in this dictionary.

The KWIC program enables one to see the context of the verb or word selected for the current paper. In this case one can see the verb *take* and the words that surround it on the left and right:

The collocations of the verb *marr/take* are extracted from the list of words:

Collocations of the verb *marr*	Use of words	Words associated with the verb *marr*	Words used without the verb *marr*
Merr menja (*I think*)	6	4	2
Marr rrugën (*take the road*)	39	2	37
Marr vendim (*take decision*)	5	3	2
Marr vesh (*understand*)	9	9	-
Marr pjesë (*participate*)	35	7	28
Marr ngrykë (*embrace*)	1	1	-
Marr parasysh (*take into account*)	4	4	-
Marr në dorë (*take care*)	17	2	15

Table 2. Collocations of the verb marr/take

The word *menja* (mind) bears the primary meaning in the phrase *merr menja* (the explanation of this collocation in the Dictionary is explained in more detail in the name (*menja*). Base is the word used in its primary and typical sense. The collocate, which is a word semantically related to the base word, is used to express meaning that differs from its usual meaning. The base in this collocation is the name *menja*, and the collocate is "marr." In terms of semantics, the name *menja* or *mind* contributes more to the meaning of *menja*, which is

"thinking". It is considered a stable or strong collocation because the noun *menja* comes four times in close proximity to the verb *marr* and inside the same sentence.Additionally, the name *rrugë* (road), which is more closely related to the meaning of the word "travel", conveys the main meaning in the example *marr rrugën*. The table below shows that the name *rrugë* is used 39 times, twice in collocation with the verb *marr*. This example constitutes a weak collocation. Another instance of this kind is *marr vendimin*, whose basic meaning is determined by the noun *vendimi* (decision). Due to its frequent appearance with the verb *marr* rather than on its own, the collocation *marr vendimin* is shown in the following table as a stable collocation. The word *vendimi* has been used in conjunction with the verb *marr* five times altogether. Even in the case of the example *marr vesh*, the definition in the dictionary is found in the noun *vesh*, not the verb *marr*, with the connotation *I understand*. However, unlike the previous example, the connection between the base and the collocate's meaning is not as obvious. This word is used as both a verb and a noun in this context (ear, dress). It appears 11 times as a noun, twice with the meaning "dress," and 9 of those times the verb *marr* precedes the noun *vesh*. This collocation constitutes another stable collocation.The meaning of the collocation *marr pjesë* can be found in the word *pjesë* (part), which denotes *I am a member or a member of something*. This is a weak collocation because the verb *marr* is used in conjunction with the term *pjesë* is used seven times out of the 35 instances.

The collocation *marr ngrykë* means embrace. In the Dictionary of the Albanian language it is found in the adverb *ngrykë*, which carries the main meaning in this construction. The adverb *ngrykë* is used once in the corpus, accompanied by the verb *marr*.In the example *marr parasysh*, the explanation is given in the adverb *parasysh (in regard)*, which means consider something. The table below demonstrates that the verb *marr* is always used before the adverb *parasysh*. So *marr parasysh* constitutes a strong collocation.In the Dictionary of the Albanian the meaning of the collocation *marr në dorë* is neither specified on the verb nor on the noun. The word *dorë* (hand) is used 17 times in the passage below, twice with the verb *marr*. Concerning the verb *marr*, this noun is also collocated with a variety of other verbs, including the verbs *lyp*/beg and *jap*/give.As can be seen from the aforementioned instances, this verb can be used with three different types of nouns, adverbs, and prepositional phrases. The three formulations are: *marr vesh, marr rrugën, marr pjesë, marr vendim, merr menja;* adverbs: *marr parasysh; marr ngrykë;* prepositional phrase: *marr në dorë.*

3.2. Collocations of the verb *bëj (do)*

In the spoken corpus of Kosovo, there are two different forms of the verb *bëj*. Along the corpus, the first form, *boj*, appears 111 times, while the other variant, *baj*, just 86 times. The first meaning of the verb *bëj* in the Dictionary of the Albanian Language is: "*prepare, produce something, by processing a subject; build or build something, using other items, different parts, etc.; cook something with work, with my hands or using a tool, a machine or another tool: I make bread. I make the dish. I build a wall, etc*". There are 39 different verb meanings of *bëj* listed in this dictionary.

Collocations of the verb *bëj*	Use of basic words:	Use of basic words with the verb *bëj*:	Use of basic words without the verb *bëj*:
Bo hyzmet - serve	2	2	-
Bo dam - damage	4	3	1
Bo gati – prepare	15	4	11
Bo marak – worry	5	2	3
Bo hallall – forgive	2	2	-
Bo vaki – happen	3	3	-
Ba për hajr - congratulate	2	2	-

Table 3. Collocations of the verb bëj/do

From the table above it can be seen that some collocations, such as: *bo hyzmet, bo hallalll, bo vaki, ba për hajër*, are stable collocations, because whenever their basic word is used, it is accompanied by the verb *bëj*, with which it forms collocations. In addition to these, the use of collocations that are less stable is noticed as well, such as: *bo marak, bo dam*, while the combination *boj gati* almost comes out as a weak collocation in this corpus. The meanings of collocations are often provided on the basic word, for example to the words *hyzmet, dam, hallal, marak, vaki, për hajr*. The basic meaning of the verb *bëj* is only present in the collocation *boj gati*. It is clear from the spoken corpus examples that the general meaning of these collocations is more closely related to the meaning of the base word, e.g. *bëj dëm- dëmtoj/harm, bëj merak-merakos/worry, bëj gati/përgatis-prepare etc*. Collocations of the verb *bëj* come out with these constructions: nouns (*hyzmet, dam, marak, hallall, vaki)*; adverb (*gati)* and prepositional phrase: (*për hajër).*

3.3. Collocations of the verb *dal*

There are 98 instances of the verb *dal* in the spoken Albanian corpus in Kosovo (representative form only). In the Dictionary of the Albanian Language, its first meaning is: *"I go from a closed place, surrounded outside it, exceeding its limits; go from the inside out; leave permanently or temporarily from a place where I have been, from a certain environment or from the ranks of a collective: Get out of the house. Get off the bus, etc."*. There is a range of 35 meanings of the verb *dal* in this Dictionary.

The most frequent collocations that the verb *dal* forms are listed in the following table along with the words it is collocated with:

Collocations of the verb *dal*	Use of the word *jashtë*	Use of the word *jashtë* with the verb *dal*	Use of the word *jashtë* without the verb *dal*
Dal jashtë – go out	36	10	26

Table 4. The collocations of the verb dal/go out

The word *jashtë*, which refers to going abroad, is frequently preceded by the verb *dal*. The word *dal* and the word *jashtë* both have higher frequency of occurrence when used separately than when they are together, making this combination of words a rather poor collocation. Even in this collocation, the adverb *jashtë* provides the explanation of the meaning.

3.4. Collocations of the verb *hyj*

The verb *hyj* in the spoken corpus is used 105 times in the *hi* variant. Its first meaning in the Dictionary of the Albanian Language is: *"I go and enter somewhere inside, walking or moving; enter a closed or fenced place. Entered secretly. "He ran in."* This dictionary gives 16 meanings of the verb *hyj*.

The words with which the verb *hyj* is most often collocated are:

Collocations of the verb *hyj*	Use of the basic words	Use of the basic words with the verb *hyj*	Use of the basic words without the verb *hyj*
Hi në luftë – go to war	13	6	7
Hi nër krah – help	3	1	2
Hi tekrar – come back	1	1	-
Hi në mode – in fashion	1	1	-

Table 5. The collocations of the verb hyj/enter

The table above demonstrates how frequently the verb *hyj* forms collocations with prepositional phrases, such as: *në luftë, nër krah, në modë*, but *hi në luftë* stands out more as a free collocation due to the word's nearly equal frequency when used alone and when combined with the verb *hyj*. The other instance, *hi nër krah*, uses the noun *krah* from common usage three times, once in conjunction with the verb *hyj*. *Tekrar* and *mod* are deemed to be consistent collocations in the following two examples, *hi tekrar* and *hi në modë*, because they are always used with the verb *hyj*. The following forms are collocated with the verb *hyj*: prepositional phrases (*në luftë, nër krah, në mode)* and adverb (*tekrar)*.

3.5. Collocations of the verb *vij*

In the spoken corpus, the verb vij (in the past tense – *kam ardhë*) is used 99 times. According to the Dictionary of the Albanian Language, its initial meaning is: *"It moves from a place towards the place where the one who speaks or hears is and I get here; I walked from here to there and ended up at the appointed place or at someone who called me; moves someone or something in my direction or gets to where I am; arrives: Friends came. "He came running".* This dictionary gives 31 meanings of the verb *vij*.

Collocations of the verb *vij* are presented in the following table:

Collocations of the verb *vij*	Use of the basic words	Use of the basic words with the verb *vij*	Use of the basic words without the verb *vij*
(ka) ardh te burri-got married	41	7	34
(ka) ardh nuse-got married	3	2	1
(ka) ardh vakti- the time has come	1	1	-
(kanë) ardh msit-	4	3	1

Table 6. The collocations of the verb vij/come

Since the core word in the first example, *ka ardhë te burri*, occurs more frequently when used alone than when combined with the collocate, it cannot be regarded as a stable collocation. However, because of the subject that is covered more in this corpus, the verb *vij* frequently collides with the prepositional phrase *te burri*. With the exception of *ka ardhë vakti*, where the word *vakt* is only used once along with the verb *vij*, these

collocations, as can be observed in the other three cases, are more related to the marriage process. The following forms are collocated with the verb *vij:* nouns (*vakti, nuse, msit*) and the prepositional phrase (*te burri*).Given all that has been said so far, it is clear that the verbs *bëj* and *marr* represented the most reliable collocations. Numerous factors, including the high frequency of occurrence of these verbs in the spoken corpus, their ambiguity (they were presented with more than 30 meanings in the dictionary), and the valence of the verb, all had an impact on the outcome. Valence is the affinity of the verb to take a complement. Only the verbs *marr* and *bëj* are ditransitive verbs. The other verbs, such as *dal, hyj*, and *vij,* are intransitive verbs. Therefore, it's possible that the verb's ability to accept complements had the effect of making the collocations between these two verbs stronger or more stable. The verb *marr* was most often used in the following contexts: *marr vesh, marr parasysh, marr ngrykë* and *merr menja*, while weaker collocations included: *marr rrugën, marr pjesë, marr në dorë*. The strongest or most stable collocations for the verb *bëj* include: *boj hyzmet, boj hallall, bo vaki, ba lazëm, ba për hajër*, while the weakest collocations were: *bo gati, bo merak, bo dam*. The verb *dal* was collocated mostly with the word *jashtë*. This word combination was deemed to be a weak collocation because the frequency of the word *dal* was higher than when used in collocation with other words. Some words appeared only once in the corpus and were regarded as stable collocations if they occurred with the verbs under investigation, for example, with the verb *hyj*: *hi tekrar, hi në modë*. The verb *vij* has the following stable collocations: (*ka*) *ardh vakti* and (*kanë*) *ardh msit*, while (*ka*) *ardh te burri* and (*ka*) *ardh nuse* were considered less stable collocations.

3.6.   Semantic type and formal aspect of collocations

Sheinfux et al. (2019) see the categorization of idioms, based on the dimension of figuration and transparency. According to them, idioms are figurative if their literal meaning can conjure up a vivid picture in the speaker's mind. In *transparent figurative idioms* the relationship between the literal picture and the idiomatic meaning is perceived to be motivated. They also distinguish *opaqe figurative idioms,* which portray a picture whose relationship to the idiomatic meaning is not perceptible. In the researched spoken corpus of Albanian in Kosovo, the second type is never found.As for the semantic type, the collocations with the verbs *take* and *do* mainly form transparent idioms, such as: *merr menja/I think, marr vesh/understand, marr ngrykë/embrace, marr parasysh/ take into account, marr në dorë/ take care*. So, in these collocations, the connection between the general meaning of the idiom and the meaning of the noun phrase or preposition phrase that is part of the collocation can be observed, for example: *mind-think, ear-understand etc*. In the other hand, the case with the collocation *go out* it's different, because this is not an idiom in the Albanian language. So, even though these two words are often used together, they do not represent idioms, because they do not give a different meaning when they are used separately as separate words. The same applies to the majority of collocations with the verb *hyj/hij/enter*, such as: *hi në luftë/go to war, hi tekrar/ come back , hi në mode/ in fashion* , which are not idioms, except for the example *hi nër krah/ help*, which in the Albanian language is usually used with the meaning *help*. The collocations of the verb *vij/come* are the same, they do not form idioms. While, regarding the formal side of these collocations, order alternation within VP and lexical insertion has been observed. While in one language there may be a fixed order of the components in the sentence, in another language the order is free. In the Albanian language, the order of the components in the sentence is free. The natural order of phrase segments in the Albanian language is NP-VP. But with an emphatic and stylistic effect, Albanian also allows inversion. According to Bokshi (2005), segments are functionally *autonomous molecules* whose value is not determined by their order, the natural order is the result of the ancient tradition of the language and not imperative (because of the need) of the structure of the language. In the collocations of the spoken corpus, it is observed that the usual order is: VP → V→ NP/ PP. In most of the examples, the subject is not present structurally, but is implied through the personal ending of the verb and through the short form of the personal pronoun used before the verb, *ma* that refers to the first person pronoun and *ta* that refers to the second person pronoun:

Nuk ma **merr menja***/ (I) don't think so.*
Ta merr **menja***/ (You) think so.*

There are cases when the complement is inserted between the collocation, as *Krejt fmija dulën me* **marr** *babën* **ngrykë***/ All the children come out to hug their father.*

Lexical insertions within the idiom, except in this case, includes adverbials, quantifiers and different types of noun modifiers. In some of the following examples, it can be seen that lexical elements have been added within the idiom:

a.   *Kishe e kish* **marr** *ni* **rrugë***/ you had taken a path.*

       b.   *Për me **marr** ndonjë **vendim**, i thirrum t'gjith anëtarët/ to make a decision, we called all the members.*

       c.   *Krejt fmija dulën me **marr** babën **ngrykë**/ All the children come out to hug their father.*

       d.   *Çajo çe ja kena **bo** çat **damin**/ to whom we caused the damage.*

In example (a) the infinitive article *ni/a* is inserted through the collocation *marr rrugën/ take the road*, in example (b) the indefinite pronoun *ndonjë/any* is inserted through the collocation *marr vendim/ take decision*, which modifies the noun *vendim/decision*, and in example (c) the complement *babën/ the father* is inserted through the collocation *marr ngrykë/ embrace*, in example (d) the demonstrative pronoun *çat/ that* is inserted, which modifies the noun *damin/damage* in the collocation *bo damin/ damage*.

In the spoken corpus, a case is presented when the meaning of a collocation is given in two different forms, as *ka ardh te burri* and *ka ardh nuse*, both of these collocations have a meaning that means she got married.

4.   Conclusions

The study of collocations can be approached from a variety of perspectives, beginning with their construction, semantics, variation, etc. This study examined the semantics and composition of the verb collocations with the highest frequency in the spoken corpus of Albanian in Kosovo. The verbs: *vij, bëj, dal, hyj, marr* and *rri* were the ones that were highlighted. The same verbs were given more than 30 meanings in the Dictionary of the Albanian Language, except for the verbs *rri* and *hyj*, which came out with 16 meanings.

The verbs *marr* and *bëj* tended to make the strongest and the most consistent collocations, for several reasons. These two verbs, in contrast to the verbs *rri, hyj, dal, and vij*, have an affinity to be followed by a complement for the reason that they are ditransitive verbs. In addition, they show high frequency of occurrence in the spoken corpus and have a range of meanings.

In stable collocations, the basic word has a stronger semantic impact than the collocate, which is semantically reliant on it.

As for the structure of collocations, the verbs were mostly combined with nouns: *marr vesh, bëj hallall, (ka) ardh vakti*; followed by adverbs: *marr ngrykë, bëj gati, dal jashtë* and finally prepositional phrases: *marr në dorë, ba për hajër, hi nër krah* etc.

REFERENCES:

Agalliu, F – Angoni, E et al. (2002)	Gramatika e gjuhës së sotme shqipe 1, ASHSH, Tiranë.
Baker, P. (2010)	Sociolinguistics and corpus linguistics, Edinburgh University Press, 24 p.
Gelbukh, A., Kolesnikova, O. (2013)	Semantic Analysis of Verbal Collocations with Lexical Functions, Springer, 2-61 p.
Lyons, J. (2001)	Hyrje në gjuhësinë teorike, Tiranë, Dituria, 408 p.
Meyer, G. (1891)	Fjalori etimologjik i gjuhës shqipe, Tiranë, Çabej.
Sinclair, J. (1991)	Corpus, Concordance, Collocation, Oxford University Press, 109-110 p.
Nesselhauf, N. (2005)	Collocations in a Learner Corpus, John Benjamin, Amsterdam.
Thomai, Jani et al. (2006)	Fjalor i Gjuhës Shqipe, Tiranë, ASHSH.
Sheinfux, Livnat Herzig et al. (2019)	Verbal multiword expressions: Idiomaticity and flexibility. Representation and parsing of multiword expressions: Current trends, Berlin: Language Science Press.
Hüning, Matthias et al. (2011)	Multi-word expressions, Berlin, De Gruyter.