Creating Lexical Phrases within Construction Grammar and their Application in Persian Dictionaries[1]

Abolfazl Alamdar[2]
Raheleh Gandomkar[3]

Abstract

Idioms and collocations are part of lexical phrases which are of high importance in lexicography. One of the criteria for distinguishing these two types of phrases from each other is the inflexibility of their constituent elements. Idioms, unlike collocations, are not flexible in terms of word order and are always used in a fixed and specific form. The purpose of the present article was to show firstly whether all Persian idioms are inflexible against syntactic changes or some are flexible; secondly, to indicate how a new lexical phrase is created as a result of converting an idiom into a collocation. To illustrate this phenomenon, a new and practical definition for idioms and collocations, based on syntactic criteria not merely semantic, was provided. According to Construction Grammar, idioms were analyzed based on four syntactic tests: 'passivization', 'clefting', 'dislocation', and 'interrogation'. Two hundred idioms were collected from two of the most up-to-date and comprehensive sources, namely *Farsi Amiyaneh Dictionary* (Abolhassan Najafi, 2008*)* and the two-volume *Dictionary of Kenayat-e Sokhan* (Hassan Anvari, 2019). Some idioms were flexible to syntactic changes, and behaved similarly to collocations. The flexibility of idioms led to introducing a new definition of these phrases based on syntactic criteria. Thus, such

[2] Ph.D. Candidate of Linguistics, Allameh Tabataba'i University, Tehran, Iran; abalfazl_alamdar@atu.ac.ir
ORCID: https://orcid.org/0000-0002-0647-9740
[3] Associate Professor of Linguistics, Allameh Tabataba'i University, Tehran, Iran (corresponding author); r.gandomkar@atu.ac.ir
ORCID: https://orcid.org/0000-0003-2281-485X

phrases were considered a kind of collocation, not an idiom. The result showed that converting an idiom into a collocation always creates a lexical phrase or lexeme with a new meaning. Finally, a plan was introduced to represent such idioms practically in Persian dictionaries.

Keywords: idiom, collocation, Construction Grammar, lexical phrase, Persian lexicography

1. Introduction

A dictionary is a reference book that provides information about the meanings of words and how to pronounce them, often showing examples of how they are used in context, and is usually presented in alphabetical order (Landau, 2001, p. 6). Authoring and compiling a dictionary is called lexicography, which is a scientific field and a specialized activity. Lexicography is a very difficult, time-consuming and challenging task. According to Atkins and Rundell (2008, p. 2), no dictionary is perfect, and almost all of them can be called as 'work in progress'; moreover, since there is not any dictionary without defects, there is no unique method for dictionary-making. These endless challenges have always made the lexicographer subject to criticism.

In active lexicography, which is used for translation from native language to foreign language, the information provided in the sub-entry is more important than the entry itself. This information often includes lexical groups or phrases such as collocations, idioms, clichés, etc. It is obvious that their accurate knowledge plays a crucial role in lexicography. One of the important distinctions between idioms and collocations has often been its semantic aspect; idiom is considered as a lexical phrase whose meaning is not the sum of the meanings of its constituent elements (*Macmillan English Dictionary for Advanced Learners*, 2007). Collocations are also a group of words that have a high co-occurrence frequency.

Another important distinction between idioms and collocations is related to the lack of syntactic changes in idioms. Therefore, one of the main characteristics of idioms is the immutability of their constituent elements. In relation to the syntactic changes of idioms, Mel'čuk points to four syntactic

changes, including the constructions of passivization, clefting, dislocation, and interrogation, which can cause the displacement of the constituent elements in idioms (2014, p. 309). In standard written and spoken Persian, we, native Farsi speakers, discover idioms in which we witness these syntactic changes. The occurrence of each of these constructions in idioms alone is enough to claim that in some idioms, we witness a syntactic change. Therefore, the discussion of syntactic invariability of the constituent elements of idioms is practically ruled out. In fact, it seems that some phrases that we know as idioms and have an entry in dictionaries indicate a syntactic behavior similar to collocations. In this article, we investigated whether these syntactic changes occur in all Persian idioms or we see changes only in a limited set; and if we observe the occurrence of this event, will our definition of idioms remain the same as the previously known definitions or will a new categorization be required in phraseology? Is it possible to divide Persian idioms into two distinct groups, fixed and flexible, based on the four tests of passivization, clefting, dislocation, and interrogation? It seems that we are not faced with an idiom in its general sense, but we will have a new form of collocation where one lexical element will play the role of a core and the other lexical element will play the role of a function for the core. In addition, it seems that Persian idioms can be divided into two categories: fixed and flexible. Therefore, the two main questions of the current research are: 1) Considering the flexibility of some idioms, what distinction can be made between these two categories of idioms and collocations? 2) How should flexible idioms be presented in Persian dictionaries?

According to the authors, changing the role of a lexical combination from an idiom to a collocation always creates a lexical group or a word with a new meaning in the vocabulary of a language. Based on this, a plan was proposed to represent idioms in which there is a possibility of these four syntactic changes in monolingual and bilingual dictionaries practically. The proposed scheme is that if a phrase accepts syntactic changes, we are handling a collocation, not an idiom. This idiom, which syntactic changes do not empty it of its original meaning, should not be the main entry in the

dictionary, but the lexical group or the word created in a new meaning shall be the main entry, so that the other words that made up the idiom will play the role of collocators in the sub-entry.

The authors have extracted 200 idioms from the most up-to-date sources, namely *Farsi Amiyaneh Dictionary* (Abolhassan Najafi, 2008) and *Kenayat Sokhan Dictionary* (Hassan Anvari, 2019). The analysis tool was the use of authors' linguistic intuition as Persian native speakers to investigate the possibility of syntactic changes. To analyze the data, sixteen Persian speakers were asked to judge the well-structured data resulting from syntactic changes.

2. Theoretical framework

This section is presented in five main subsections. In section 2.1., first, the Construction Grammar and the most important purposes of this grammar, as the theoretical framework of the present research, are presented. Then, the four constructions used in this research (i.e. 'passivization', 'clefting', 'dislocation', and 'interrogation') are mentioned.

2.1. Construction Grammar

"Construction Grammar (CxG)" is a cover term for a number of grammatical theories and models in cognitive linguistics. The most basic unit of study in this model is "grammatical construction", not syntactic units or those rules that combine these syntactic units with each other. According to the CxG, the basis of establishing linguistic communication is a set of fixed expressions in the language speakers' minds, which are committed to the mind like a formula. From this point of view, linguistic expressions as pairs of form and meaning are considered as the main and formal unit of language (Goldberg, 1995, 2003; Sinclair, 2004).

In the CxG, each construction to some extent has a degree of conceptual and abstract content that has various stylistic representations. Constructions in this grammar, just like lexemes, are a combination of form and meaning. These constructions are basically considered symbolic units

(Croft, 2007, p. 72). These constructions are associated with syntactic, morphological, phonological, and pragmatic meaning. According to Goldberg, C is considered a construction if and only if C is a pair of form and meaning <Fi, Si> in such a way that some aspects of Fi or some aspects of Si cannot be completely predicted from the components of the C construction. (1995, p. 4). In this definition, F stands for the word form and S stands for the word semantics. Therefore, the pair <F, S> represents a symbolic unit. Indexes demonstrate the symbolic relationship between form and meaning. As a result, constructions can be considered the intersection of syntax and semantics. Goldberg believes constructions are "mental patterns and frames' that can be 'morpheme', 'word', 'idiom', 'verb phrase', 'noun phrase', 'ditransitive verb' and 'complex word'" (2003, p. 220).

Based on CxG, grammatical knowledge is uniformly represented in language speakers' mind, which are present as generalized constructions. All linguistic units, from words to general syntactic and semantic rules, can be explained in the form of construction. In fact, CxG generalized construction to all grammatical constructions comprising form and meaning. In addition, the examination of idiomatic constructions from another angle, within the framework of syntactic theory, led to the emergence of a uniform representation of all grammatical constructions.

One of the most important features of "construction" is that all its meanings cannot be derived from the meanings of its constituent elements. These constituent elements can be words or morphemes or even phrases (Goldberg, 1995). It is worth reminding that constructions can appear in different sizes and levels of complexity, from a complete sentence to various phrases as well as word-formation patterns.

The main hypothesis in Goldberg's Construction Grammar is that the constructions themselves have meaning in the sentence, and this meaningfulness of the constructions is separate from the meaning of the words used in the sentence (Goldberg, 1995, p. 1). It seems necessary to mention that Goldberg never denies the fact that words play a major role in conveying meaning, but she believes that a "lexical grammatical model" alone

cannot fully and accurately represent the meaning of a sentence. As mentioned at the beginning of this section, a construction in CxG is a symbolic unit which can be represented as follows:

Figure 1

The symbolic structure of construction in Construction Grammar (Croft, 2001. p. 18)
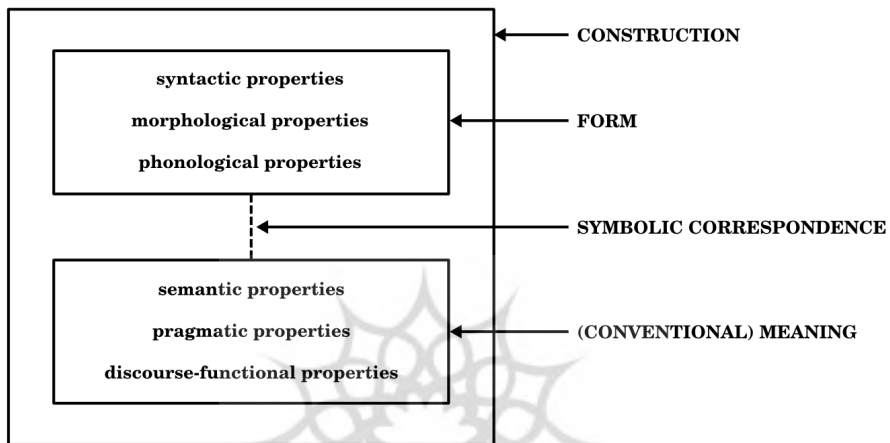
For the first time, Newmeyer (1974) mentioned the possibility of applying four syntactic tests (proposed in the introduction section) to idioms. Then, Schenk (1995) applied the same four tests to the idiom "to kick the bucket". The main goal of the present research is to examine the Persian idioms within these four constructions. Before analyzing the idioms, it is necessary to briefly introduce these four syntactic tests to clearly determine the constructions in which the idioms will be examined.

2.2. Passivization

The conversion of active construction into passive construction is done based on three principles, which are actually three universal processes in the world's languages (Perlmutter & Postal, 1977):

1- The direct object of the active sentence replaces subject in the passive sentence (Perlmutter & Postal, 1977, p. 399). For example, in the active sentence "Abbas killed all the enemies", the phrase "all the enemies" is the direct object which plays the role of the subject in its passive sentence "All the

enemies were killed".

2- When the subject of an active sentence appears in the passive sentence, it is a *chômeur element* (Perlmutter & Postal, 1977). For example, in the passive sentence "All the enemies were killed by Abbas", the word "Abbas" is a *chômeur element*.

3- The passive sentence in surface structure is an intransitive sentence (Perlmutter & Postal, 1977). For example, the passive sentence "The tire was punctured" is an intransitive sentence.

According to Dabir-Moghaddam (2020, p. 43), after applying the passive transformation on the underlying structure of active sentences, passive sentences are obtained which are unambiguous and they only have a passive concept. He argues that the passive transformation cannot be applied to all Persian verbs. Therefore, the passive construction in Persian language is transformation-bound and is applied only in those verbs that have a special semantic feature (Dabir-Moghaddam, 2020). For example, sentence (1) is an active sentence and sentence (2) is its passive equivalent.

(1) Iranian football fans consider Ali Daei as the best football player in the history of their country.

(2) Ali Daei is considered the best football player in the history of his country.

2.3. Clefting

The term "cleft sentence" was used for the first time by Jespersen (1909). The equivalent of Jespersen's example can be seen in Persian language as a sentence like sentence (3).

(3) [It]pronoun was Mehran [who ate the food]relative clause.

This sentence consists of the following parts:

1- Pronoun 'it' which is called cleft pronoun.

2- The noun after the pronoun, which is called focus or cleft constituent. In example (3), the proper noun Mehran is the focus or cleft constituent.

3- Copula

4- A type of relative clause. Some researchers, such as Quirk et al. (1985), Sornicola (1988), and Miller (1999) do not consider this type of clause as a relative clause.

2.4. Dislocation

One of the marked syntactic constructions in Persian language, in terms of word order, is dislocation (Moezzipoor, 2018, p. 46). A dislocation is a construction in which a constituent that can be an adjunct or an argument of a clause appears outside the clause, either on the right or on the left side (Sanchez & Ott, 2020). The dislocation on the left side is called left-dislocation. In this construction, the oblique object or possessor in an *ezafe* construction leaves its original position within the clause and gives its position to a pronominal clitic that carries the syntactic role of the left-dislocation constituent. In the Persian language, this dislocation is usually followed by the postposition 'rā' (Moezzipoor, 2018, p. 46). He gives example (4) for possessive left-dislocation in an *ezafe* construction, and example (5) for the left-dislocation of the oblique object[1]:

(4) *qazal-o[2]*   *šomār-aš-o*   *peydā*   *kardam*

     Qazal-DO   number-her-DO   apparent   did-I

     'As for Qazal's number, I found it'.

(5) *bačče-ro*   *čāqu-ro*   *az-aš*   *begir*

     child-DO   knife-DO   from-his or her   take

     'As for the child, take the knife from him or her'.

---

[1] In this article, the following notations and abbreviations were used:
[a]: low front vowel
[ā]: low back vowel
pl: plural
part: participle
EZ: Ezafe (genitive) marker
DO: Direct Object
NEG: Negative
INF: Infinitive marker
HIATUS: a phoneme which prevents two vowels from appearing next to each other.
Ø: Zero morpheme which indicates verbal agreement with third person singular subject when the verb is used in the past tense.
[2] -o is a spoken form of the postposition rā (the direct object marker in Persian). Another spoken form of rā is ro.

In the dislocation construction on the right side, a noun phrase is moved to the right side of the clause, and the clause itself puts a coreferential pronoun instead of that moved noun phrase (Toma, 2018, p. 1). The dislocation on the right side is called *right-dislocation*. Ward & Birner (1996, p. 472) provide example (6) for the right-dislocation construction:

(6) She$_i$'s a smart cookie, that Diana$_i$.

2.5. Interrogation

In Chomsky's Universal Grammar, movement is one of the most important syntactic phenomena. One construction in which the phenomenon of movement is investigated is the interrogative construction. Movement consists of rules called transformation. In the movement process, the linguistic constituent moves from its position and occupies another position. Chomsky calls this event α movement. The movement of wh-expression is one of these movements that has attracted a lot of attention. According to Chomsky, wh-expression is moved from its original position to a position at the beginning of the sentence, which is a non-role position. In other words, there is a movement from subject to non-subject position. He calls this secondary position the *complementizer phrase specifier*. English is one of the languages in which the movement of the wh-expression is mandatory, but there are also examples where the wh-expression does not move, like (7) that Radford (2009, p. 183) brings as an echo-question:

(7) She was meeting who?

Regarding such examples, Chomsky considers movement in English to be of two types:

A) Overt syntactic movement. In this movement, the wh-expression is obviously transferred to the position of the complementizer phrase specifier and a grammatical sentence is obtained.

b) Non-overt syntactic movement. The wh-expression remains in its original position (logical level) and does not reach the phonetic level. In the above example, a non-overt syntactic movement is seen.
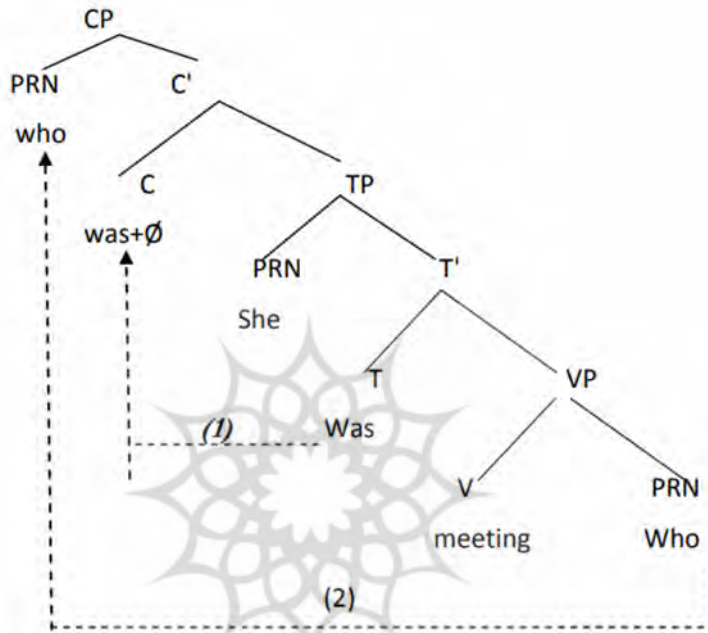
To illustrate the overt syntactic movement, Radford gives example (8)

as well as the following figure (2009, p. 183-185).

(8) Who was she meeting?

Figure 2

Obvious Syntactic Movement (Radford, 2009, p. 184)

As it is clear in figure 2, the wh-expression 'who' moves from the position of verb phrase to the position of the pronoun in the complementizer phrase. The verb 'was' moves from the position of the tense phrase head to the position of the complementizer phrase head, as well.

3. Method

In the current research, the authors have tried to give, based on the syntactic changes, a new definition of Persian idioms, and then explain how a lexical group is created, resulting from this phenomenon. For this purpose, it has been tried to use the most up-to-date sources available. Researchers have collected the idioms from two main sources, namely *Farsi Amiyaneh Dictionary* (Abolhassan Najafi, 2008) and *Kenayat Sokhan Dictionary* (Hassan Anvari, 2019).

The data were analyzed based on the authors' linguistic intuition as Persian speakers to investigate the possibility of syntactic changes. Then, the idioms were listed according to the previously mentioned definition. In the next step, the changes and displacements of the elements present in idioms were examined, then, sixteen Persian speakers were asked to judge the well-structured data resulting from syntactic changes. In addition, it was indicated how a lexical phrase resulting from the above-mentioned syntactic tests, i.e. passivization, clefting, dislocation, and interrogation, was created. Finally, its application in Persian lexicography was presented.

4. Discussion

Since the mentioned syntactic constructions are marked in Persian language, they have always been of great interests to researchers; however, limited research has been done on the syntactic constructions of idioms, and so far a comprehensive definition of idioms and collocations has not been proposed based on the probability of occurrence of these constructions in idiomatic expressions. In addition, the possibility of these constructions occurring in Persian idioms has been a question that has not been answered before this research.

In this section, it is tried, while analyzing 200 Persian idioms and applying the principles and concepts raised in the Goldberg's CxG framework, to provide an accurate and comprehensive definition of idioms based on the syntactic features of this type of expressions. By examining these data, it is determined which idioms accept some or all syntactic constructions (passivization, clefting, dislocation and interrogation) and which types of idioms do not show the flexibility. In addition, the question "what is the reason for the inflexibility of some idioms?" will be answered.

After performing these analyses, we will discuss how a lexical phrase or word in a new and limited meaning is created, and finally, the way of representing these phrases or words in Persian lexicography is shown.

4.1. Data analysis

In this section, 24 of the 200 selected idioms are presented as

examples in the following four subsections. Then, they are analyzed, and the flexible ones will be determined. The flexibility of each idiom is indicated by a tick mark (✓). Using this sign in front of the idiom means that the phrase in question does not lose its idiomatic meaning. For example, in the case of the expression *havās-e kasi-rā part kard-an* (=to distract somebody's attention), a sentence like *havās-am part šod* (=I was distracted) can be used in the passive construction. In this sentence, the same meaning as the original phrase is taken from it. As a result, the above idiom is flexible against the passive construction and is indicated by a tick mark. The inflexibility of an idiom is showed by an asterisk mark (*), which indicates that the sentence is meaningless, or at least has a less established meaning among Persian speakers. It is worth noting that all the sentences that are made by applying the four syntactic tests on idioms have a meaning, but that idiomatic meaning may not be taken. For example, in the idiom *hendevāne zir-e baqal-e kasi gozāšt-an* (=to flatter or praise somebody; literally: to put a watermelon under somebody's armpit), its passive construction is considered being asterisked and therefore meaningless. In the sentence *zir-e baqal-e ali (az su-y-e pedar-aš) hendevāne gozāšt-e šod* (literally: A watermelon was put under Ali's armpit (by his father)), which is the passive form of the above idiom, although the sentence has a meaning (for example, Ali's father has bought a watermelon and put it under his son's armpit to take it home), the idiomatic meaning (to flatter somebody) cannot be obtained. Some Persian idioms cannot be used in the passive or clefting constructions. For example, the idiom *morq-e kasi yek pā dāšt-an* (=to be persistent and not to go back on one's word), and also the idiom *rudarvāsi-rā kenār gozāšt-an* (= to express one's intention clearly) do not exist in the passive and clefting forms, respectively. In these cases, the flexibility of such constructions is indicated by a dash mark (-). The reason for this problem is explained in the subsections related to the passive and clefting.

4.2. Passive construction in idioms

The first construction examined in relation to idioms is the passive construction. Initially, we evaluate the flexibility of idioms in the passive

construction in the form of 6 idioms. Here, there are three categories of idioms in the passive constructions as follows:

(9) *jegar-e   kasi-rā   suzand-an*

 liver-EZ   somebody-DO   burn-INF

 literal meaning: to burn somebody's liver;

 idiomatic meaning: to make somebody suffer deeply.

Passive construction:

 jegar-e   mādar-aš   suxt-Ø

 liver-EZ   mother-his or her   burnt-it

✓ 'His or her mother's liver burned'.

(10) *ātaš   be   pā   kard-an*

 fire   to   foot   do-INF

 literal meaning: to set fire;

 idiomatic meaning: to cause chaos.

Passive construction:

 ātaš   be   pa   šod-Ø

 fire   to   foot   became-it

✓ 'The fire was set'.

The first category shows flexibility against the passive construction. In other words, the construction under consideration does not empty the idiom from its idiomatic meaning. In Persian, unlike in English, the passive construction does not significantly change the constituent elements of an idiom, and it is done simply by converting one light verb into another. According to Karimi (1997), more than 90% of Persian verbs are made with light verbs. A high percentage of Persian light verbs are also made with the light verb *kardan* (to do), to the extent that this verb is also often used with foreign words, such as *email kardan* (to send an email, to email), *kāt kardan* (to cut), *kansel kardan* (to cancel), etc. By all means, this does not mean that whenever an idiom contains the verb *kardan* or other light verbs, it can be converted into passive.

(11) *ārezu   be   del-e   kasi   mānd-an*

 desire   to   heart-EZ   somebody   stay-INF

 literal meaning: a desire to stay in somebody's heart;

idiomatic meaning: not to fulfill somebody's wish and always miss it.

Passive construction: - (An intransitive verb)

(12) *morq-e   kasi   yek   pā   dāsht-an*

        hen-EZ   somebody   one   foot   have-INF

        literal meaning: somebody's hen to have one leg;

        idiomatic meaning: to be persistent and not to go back on one's word.

        Passive construction: - (A transitive verb that cannot be passivized)

        The second category is idioms which can't come in passive form. For example, idioms such as *ārezu be del-e kasi mānd-an* or *morq-e kasi yek pā dāsht-an*, can't be converted into passive, since an intransitive verb has been used in the structure of these expressions.

(13) *hendevāne   zir-e   baqal-e   kasi   gozāšt-an*

        watermelon   under-EZ   armpit-EZ   somebody   put-INF

        literal meaning: to put a watermelon under somebody's armpit;

        idiomatic meaning: to flatter or praise somebody.

Passive construction:

        zir-e   baqal-e   ali   hendevāne   gozāšt-e   šod-Ø

        under-EZ   armpit-EZ   Ali   watermelon   put-part   became-it

* 'A watermelon was put under Ali's armpit'.

(14) *šotor-rā   bā   namad   dāq   kard-an*

        camel-DO   with   felt   hot   do-INF

        literal meaning: to heat the camel with felt;

        idiomatic meaning: to make the work very difficult for a person.

Passive construction:

        šotor   bā   namad   dāq   šod-Ø

        camel   with   felt   hot   became-it

* 'The camel was heated with felt'.

        The third category is idioms such as *šotor-rā bā namad dāq kard-an* which its passive form *šotor bā namad dāq šod-Ø* (=the camel was heated with felt) cannot be used in the same idiomatic sense. One of the reasons is that the expression 'the camel was heated with felt' is not used as an idiom among Persian speakers, and this has caused the expression *šotor-rā bā namad dāq*

kard-an to maintain its high level of inflexibility.

4.3. Clefting construction in idioms

The next construction in idioms is clefting construction. We examine the behavior of 6 idioms in the cleft construction and then provide a model of the idioms' flexibility or inflexibility in the construction under consideration. Here, we face two categories of idioms in clefting constructions as follows:

(15) *del-e   kasi-rā   šekast-an*

heart-EZ   somebody-DO   break-INF

literal meaning: to break somebody's heart;

idiomatic meaning: to turn somebody's hope into disappointment.

Clefting construction:

in   del-e ali bud ke   šekast-i

this   heart-EZ Ali was   that   broke-you

✓ 'It was Ali's heart that you broke'.

(16) *havās-e   kasi-rā   part   kard-an*

Senses-EZ   somebody-DO   outlying   do-INF

literal meaning: to throw somebody's senses;

idiomatic meaning: to distract somebody's attention.

Clefting construction:

in   havās-e man bud   ke   part   šod-Ø   na   to

this   senses-EZ I   was   that   outlying   became-it   NEG   you

✓ 'It was my attention that got distracted, not yours'.

(17) *ajal-e   kasi   resid-an*

life deadline-EZ   somebody   arrive-INF

literal meaning: to arrive somebody's life deadline;

idiomatic meaning: to be the time of somebody's death.

Clefting construction:

In   ajal   bud   ke   āxar-eš   az   rāh   resid-Ø

this   life deadline   was   that   end-its   from   way   arrived-it

✓ 'It was the life deadline that finally arrived'.

The first category shows flexibility against clefting construction. That

is, this construction does not empty the idiom from its idiomatic meaning. To be more precise, these idioms have a high frequency among Persian speakers in their clefting construction, and replacing their constituent elements does not make the idiom empty of its idiomatic meaning. In the following, after presenting the next three examples, we will discuss the reason for this phenomenon.

(18) *sar-e   gāv   tu-y-e   xomre   gir   kard-an*

     head-EZ cow   inside-HIATUS-EZ   vat   trapped   do-INF

     literal meaning: to get the cow's head trapped in the vat;

     idiomatic meaning: to face an unexpected problem that cannot be solved.

Clefting construction:

     in   sar-e   gāv   bud   ke   tu-y-e   xomre   gir   kard-Ø

     this   head-EZ cow   was   that   inside-HIATUS-EZ   vat   trapped

did-it

* 'It was the cow's head that got trapped in the vat'.

(19) *riq   az   damāq-e   kasi   birun   āmad-an*

     thin stool from   nose-EZ somebody outside came-INF

     literal meaning: to come out thin stool of somebody's nose;

     idiomatic meaning: to die.

Clefting construction:

     in   riq   bud   ke   az   damāq-e   u   birun   āmad-Ø

     this   thin stool   was   that   from   nose-EZ   he or she   outside

came-it

* 'It was the thin stool that came out of his nose'.

(20) *tasme   az   gorde-y-e   kasi   kašid-an*

     belt   from   kidney-HIATUS-EZ   somebody   pull-INF

     literal meaning: to pull a belt from somebody's kidney;

     idiomatic meaning: to intimidate somebody and make them submit to their will.

Clefting construction:

     in   tasme   bud   ke   az   gorde-y-e   mā   kašid-and

     this   belt   was   that   from   kidney-HIATUS-EZ   we   pulled-they

* 'It was the belt that they pulled from our kidney'.

The second category is idioms in which it is not possible to have clefting constructions. The reason for the difference is their constituent elements and the role of these elements in assuming the semantic load of the idiom. In the first category, in which the cleft construction still preserves the idiomatic meaning, there are elements that contain all or a large part of the idiom's meaning. For example, in the first to third idioms, that is, *del-e kasi-rā šekast-an, havās-e kasi-rā part kard-an*, and *ajal-e kasi resid-an*, the lexemes *"del"*, "*havās*" and *"ajal"* alone can be used in the meanings close to 'feeling or the center of feelings and emotions', 'mind' and 'time of death' respectively. This relative independence of words and their use (in the same meaning as mentioned) in other sentences, causes these words to free themselves from their absolute constraints in these idioms and gradually lower the idiom's inflexibility level. This will eventually lead to the transformation of an idiom into a collocation. In other words, when a word carries the main meaning of an idiom, it obtains the power to change its place within the structure of the idiom. When a word cannot freely change its place within the structure of the idiom, it is because it does not play a role in determining the idiom's meaning. The smaller the role of a word in determining the idiom's meaning, the less freedom of action it will have in freeing the constraints of the idiom. This applies to the second category. For example, in the expression *riq az damāq-e kasi birun āmad-an* (= to die), none of the constituent elements of the idiom have a role in determining the overall meaning of the idiom. Neither "*riq*", nor *"damāq"* nor *"birun āmad-an"* have anything to do with the concept of 'dying'. One of the important factors in the transformation of an idiom into a collocation is the commonness of that idiom among people, and especially among young people. The wide spread of an idiom causes its constituent elements to find an independent identity over time; as a result, they carry the idiom's semantic load. For example, it seems that the idiom *gir dad-ān* (= to ask somebody for something with persistent insistence) became popular among people when the TV series *Zir-e Asemān-e Šahr* (literally. Under the City's Sky) started airing in 2003. After some time, the noun *"gir"* became free from its light verb and could

be used alone. More precisely, the light verb *dad-ān* (= to give) was reduced from the sentence level (Safavi, 2012) and the word *"gir"* was used alone in many sentences among people.

4.4. Dislocation construction in idioms

The third construction examined in the present study is called dislocation. In this section, we examine the behavior of idioms in this construction and after that, present a model of the idioms' flexibility or inflexibility in the construction under consideration. Here, we face three categories of idioms in the dislocation construction as follows:

(21)   *jour-e   čizi-rā   kešid-an*

tyranny-EZ   something-DO   pull-INF

literal meaning: to pull the tyranny of something;

idiomatic meaning: to face the punishment of one's bad action.

Dislocation construction:

tanbali   jour-eš-o   hālā   xāh-ad   kešid[1]-Ø

laziness   tyranny-its-DO   now   will-he or she   pulled

✓ 'Laziness, he or she tolerates its punishment now'.

(22)   *ārezu-y-e   čizi-rā   be   gur   bord-an*

desire-HIATUS-EZ   something-DO   to   grave   carry-INF

literal meaning: to carry the desire for something to the grave;

idiomatic meaning: not to reach one's desire and die.

Dislocation construction:

demokrāsi   ārezu-y-aš-rā   be   gur   xāh-and   bord

democracy   desire-HIATUS-its-DO   to   grave   will-they   carried

✓ 'Democracy, they will carry their wish to the grave'.

The first category demonstrates flexibility against this construction. That is, the construction does not empty the idiom from its idiomatic meaning. But the important point is that the dislocation construction works only on those idioms that have lexical units such as 'something' or 'somebody'. In fact, in such

---

[1] In Persian, the future tense auxiliary agrees with the subject, and the main verb appears in its past stem form without a subject agreement marker (Dabir-Moghaddam, 2020, p. 118).

idioms, we see the *ezafe* (genetive) combination of noun + noun (for example, *ārezu-y-e demokrāsi* in the second case) where the second element can be moved to the beginning of the sentence. The reason for this is that the second element can be any word, for example *ārezu-y-e piruzi* (a desire for victory), *ārezu-y-e āzādi* (a desire for freedom), *ārezu-y-e qabul šod-an dar konkur* (a desire to be accepted in university entrance exam), etc., and it is not a fixed element of the idiom. This lowers the inflexibility level of the idiom and makes it more like a collocation than an idiom. According to the authors, the existence of such lexical units such as 'somebody' or 'something' makes a phrase very close to a collocation.

(23) *āsemān rismān be ham bāft-an*

      sky string to each other weave-INF

      literal meaning: to weave the sky (and) string together;

      idiomatic meaning: to speak scattered and irrelevant words.

Dislocatipn construction: -

(24) *band-rā āb dād-an*

      Dam-DO water give-INF

      literal meaning: to water the dam;

      idiomatic meaning: to say what should not be said.

Dislocation construction: -

      The second category is idioms in which it is not possible to make them dislocated. Dislocation can be seen only when there is an *ezafe* combination in the phrase. In this second category there is no *ezafe* combination of noun + noun. Therefore, the formation of dislocation construction cannot occur in them.

(25) *yek rude-y-e rāst tu-y-e čekam-e kasi na-bud-an*

      one intestine-HIATUS-EZ right inside-HIATUS-EZ stomach-EZ somebody NEG-be-INF

      literal meaning: not to be a right intestine in sb's stomach;

      idiomatic meaning: to be a liar.

Dislocation construction:

      čekam-e farbod yek rude-y-e rāst tu-š na-bud-Ø

stomach-EZ   Farbod   one   intestine-HIATUS-EZ   right   inside-its
NEG-was-it

* 'Farbod's stomach, there is not a right intestine in it'.

(26) *āb   tu-y-e   guš-e   kasi   kard-an*

water   inside-HIATUS-EZ   ear-EZ   somebody   do-INF

literal meaning: to pour water into somebody's ear;

idiomatic meaning: to deceive somebody.

Dislocation construction:

guš-e   mehān   tu-š   *āb   rixt-and*

ear-EZ   Mehran   into-its   water   poured-they

* 'Mehran's ear, they poured water into it'.

Finally, the third category consists idioms in which the presence of dislocation construction makes the phrase void of its idiomatic meaning and as a result, they are displayed with an asterisk. For example, in the expression *yek rude-y-e rāst tu-y-e čekam-e kasi na-bud-an* (= to be a liar), no lexical element (except for the word '*kasi*' (somebody)) can be moved to the beginning of the sentence or anywhere else. This is because in this idiom, the meaning of 'being a liar' is distributed among all the elements of the expression, and there is no element in the idiom that can alone carry all or most of the idiom's meaning. In other words, the semantic value of all elements is the same. On the opposite point, we have idioms like *jour-e čizi-rā kešid-an* where a single lexical unit (here, "*jour*") carries most of the semantic load of the idiom alone. In this idiom, "*jour*"can be considered equivalent to 'punishment of a bad deed'; while in an expression like *yek rude-y-e rāst tu-y-e čekam-e kasi na-bud-an*, there is no such a lexical unit that carries most of the semantic load of the idiom alone.

4.5. Interrogation construction in idioms

The last construction examined in the present study is the interrogation construction. First, we examine the behavior of six idioms in the construction under consideration. After that, a model of idiom's flexibility or inflexibility in this construction is provided. Here, we are faced with two categories of idioms in the interrogation construction as follows:

(27) *šāx-e   qul-rā   šekast-an*

   horn-EZ   monster-DO   break-INF

   literal meaning: to break the monster's horn;

   idiomatic meaning: to do a superhuman task.

Interrogation construction:

   fekr   kard-i kodām   šāx-e   qul-rā   šekast-i

   thought   did-you   which   horn-EZ   monster-DO   broke-you

✓ 'Which monster's horn did you think you broke'?

(28) *qāl-e   qaziye   kand-e   šod-an*

   din-EZ   matter   come off-part   become-INF

   literal meaning: to come off the din of a matter;

   idiomatic meaning: to put an end to.

Interrogation construction:

   qāl-e   kodām   qaziye   kand-e   šod-Ø

   din-EZ   which   matter   come off-part   became-it

✓ 'Which din of the matter was come off'?

(29) *(kasi-rā)   tu-y-e   ātaš andāxt-an*

   (somebody-DO)   inside-HIATUS-EZ   fire   throw-INF

   literal meaning: to throw (somebody) into the fire;

   idiomatic meaning: to cause somebody's misfortune.

Interrogation construction:

   to-rā   tu-y-e   kodām   ātaš andāxt-am ke in   qadr asabāni šod-i

   you-DO   inside-HIATUS-EZ   which   fire   threw-I   that   this much   angry   became-you

✓ 'In which fire did I throw you that you got so angry'?

   The first category shows flexibility against the interrogation construction. That is, this construction does not empty the idiom from its idiomatic meaning. In this category, that is the interrogative construction which still preserves the meaning of the idiom, the reason for that is the existence of elements that contain all or a large part of the idiom's meaning. For example, in the first three idioms, that are *šāx-e qul-rā šekast-an*, *qāl-e qaziye kand-e šod-an*,

and *kasi-rā tu-y-e ātaš andāxt-an*, the words or lexical phrases *šāx-e qul, qaziye,* and *ātaš* alone can be respectively used to the meanings close to "extraordinary work", "the flow of something" and "misfortune". This relative independence of words and their use (in the meanings as mentioned) in other sentences causes these words to free themselves from their absolute constraints in these idioms and gradually lower the level of idiom's inflexibility. This will lead to the transformation of an idiom into a collocation. In other words, when a word carries the main meaning of an idiom, it obtains the power to change its place within the idiom's structure. When the word cannot freely change its place within the idiom's structure, it is because it does not play a role in determining the meaning of the idiom.

(30) *šāx*   *tu-y-e*   *jib-e*   *kasi*   *gozāšt-an*

      horn   inside-HIATUS-EZ   pocket-e   somebody   put-INF

      literal meaning: to put a horn in somebody's pocket;

      idiomatic meaning: to encourage somebody to do something by flattery and trickery.

Interrogation construction:

      šāx-r ā   *tu-y-e*   *kodām*   *jib-e*   *mehrān*   *gozāšt-i*

      horn-DO   inside-HIATUS-EZ   which   pocket-EZ   Mehran   put-you

*＊'In which Mehran's pocket did you put the horn'?

(31) *fil-e*   *kasi*   *yād-e*   *hendustān*   *kard-an*

      elephant-EZ   somebody   remembrance-EZ   India   do-INF

      literal meaning: to make somebody's elephant remember India;

      idiomatic meaning: to think of one' halcyon days.

Interrogation construction:

      kodām   *fil-e*   *ali*   *yād-e*   *hendustān*   *kard-Ø*

      which   elephant-EZ   Ali   remembrance-EZ   India   did-he

*＊'Which Ali's elephant remembered India'?

(32)   *qāti-y-e*   *morq-hā*   *šod-an*

      blended-HIATUS-EZ   hen-pl   become-INF

      literal meaning: to be blended in with hens;

      idiomatic meaning: to marry.

Interrogation construction:

u	*qāti-y-e*	*kodām*	*morq-hā*	*šod-Ø*
he or she	blended-HIATUS-EZ	which	hen-pl	became-he or she

* 'Which hens was he blended in'?

The second category refers to the idioms in which it is not possible to make an idiom interrogative. The reason for the difference in the behavior of these two types of idioms is their constituent elements and the role of these elements in assuming the idiom's semantic load. The less the role of the word in determining the idiom's meaning, the less freedom of action it will have in freeing the constraints of the idiom. This applies to this second category. For example, in the idiom *qāti-y-e morq-hā* š*od-an* (= to marry), none of the constituent elements of the idiom plays a role in determining the overall meaning. Neither *"morq-hā"*, nor *"qāti šod-an"* have anything to do with the concept of "to marry". It is important to mention that in this section, only those idioms where examined which the question words such as *"če"* (what) or *"kodām"* (which) is used in the structure of the idiom, not those idioms where the question word appears at the beginning of the idiom. That's why the researchers were to investigate the formation change of the idioms. The question words used at the beginning of the idiom do not change the structure of idioms; therefore, they are not a suitable benchmark to evaluate the formation change of the idioms. Of course, in Persian language, question words can be used at the beginning of the sentence without changing the idiom's structure. For example, the idiom *āb az galu-y-e kasi payin na-raft-an* can be interrogated as *kodām āb az galu-y-e man payin na-raft?*, but this interrogative sentence does not specify exactly whether the idiom is flexible to the interrogation construction or not.

4.5. Data statistical analysis

Out of the total of 200 idioms that were examined, 38 idioms were used in the passive construction and produced sentences that had the same idiomatic meanings. 48, 61, and 50 idioms also produced meaningful sentences in clefting, dislocation, and interrogation constructions, respectively. The result of

the final analysis of these 200 idioms indicates that 24.62 percent of the idioms show flexibility against these four syntactic constructions. The above results prove that the constituent elements of some idioms can be moved and used in different constructions like collocations. Figure 1 shows the frequency of flexibility of idioms in four syntactic constructions.

Figure 1

Frequency of flexibility of idioms in four syntactic constructions
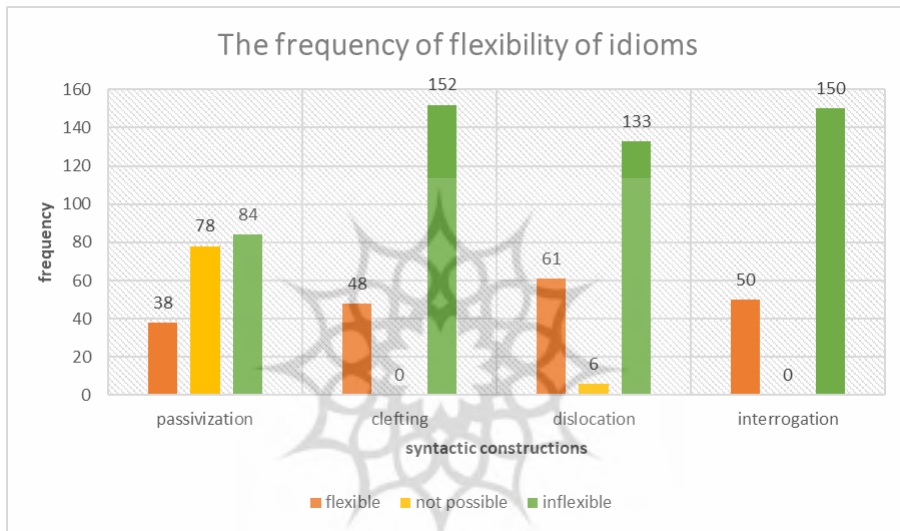
5. Conclusion

The main goal of the present research was to analyze idioms with a novel perspective and provide a new classification of these limited lexical phrases. In fact, by using the above four syntactic tests, we tried to identify two types of idioms in Persian. In addition to the main goal, another goal was followed, which is related to the presentation of idioms in dictionaries. Almost all monolingual and bilingual dictionaries present idioms as a main entry, which makes the user's hands tied when translating from his/her mother tongue to the foreign language, and he/she cannot use these idioms in different ways in his/her text as they want. If syntactic changes are possible in idioms, these apparently immutable lexical phrases will not be displayed as the main entry. In fact, they will be divided into a core and one or more functions. The

idiom should be considered as a sub-entry of the core so that the user can use that expression in a desired form, such as passive, dislocated, interrogative, etc. in his/her text.

It seems that with the increase of creativity in using language, especially among young people, we will see more and more idioms converting into collocations. Of course, the fewer the constituent elements of an idiom, the more likely it is to become a collocation. For example, the phrase *del-e kasi rā šekast-an* (= to turn somebody's hope into disappointment; literally: to break somebody's heart) consists of two lexical elements: 1. *"del"* (= heart), 2. *"šekast-an"* (= to break). However, the phrase *band-e del-e kasi pare šod-an* (= to become extremely distressed; literally: to be torn somebody's heartstrings) consists of three elements: 1. *"band"* (= string), 2. *"del"* (= heart), 3. *"pāre šod-an"* (= to be torn). This difference in the number of idiom elements makes the second idiom, unlike the first one, remain still an idiom. This is due to the fact that when the number of constituent elements is high (more than two elements), no elements can carry the main burden of meaning alone and weaken the rest of the elements in terms of meaning. For example, in the idiom *pust-e xarboze zir-e pā-y-e kasi andāxt-an* (= throw a melon skin under somebody's feet, literally: to deceive), the meaning of 'to deceive' is equally distributed over all constituent elements, and no element can assume the meaning of 'to deceive'. As a result, it is very unlikely that this idiom will be converted into a collocation in the future. But in a phrase such as *āh dar basāt na-dašt-an* (= to be clean and penniless), the lexeme *"āh"* can alone carry the meaning of the phrase (in a meaning such as lack of money). Or in a phrase like *ātaš be pā kard-an* (to cause chaos), since it has only two constituent elements (*"ātaš"* and *be "pā kard-an"*), *"ātaš"* assumes the meaning of the phrase (in a meaning like chaos) and the verb *"be pā kard-an"*, which has given up the main semantic load, becomes a light verb and plays the function role for the pivot (here *"ātaš"*) in various syntactic constructions.

According to the authors, changing the role of a lexical combination from an idiom to a collocation always results in the addition of a lexical group or a word with a new meaning to the lexicon of a language. To make it clearer,

let us propose the famous idiom 'to kick the bucket' in English. If one day all four syntactic changes mentioned in section 1 were meaningful, then, according to the authors, the lexical group 'the bucket' with the meaning of 'death', 'passing' or similar words close to the concept of death would be added to the lexicon. As a result, 'the bucket' should be presented as the main entry in dictionaries, and the verb 'kick' will act as a collocate of the lexical group 'the bucket'. In this case, the verb 'kick' in the idiom 'to kick the bucket' is a light verb and has the same role as the light verb *"kardan"* (to do) in the collocations such as *"eštebāh kardan"* (= to make a mistake), *"ezdevāj kardan"* (= to marry) or *"guš kardan"* (= to listen) because all these three verbs imply doing something.

Now we will examine the same issue in Persian. As mentioned earlier, almost all monolingual and bilingual dictionaries display idioms as the main entry. For example, an idiom such as *pust-e xarboze zir-e pā-y-e kasi andāxt-an* (= throw a melon skin under somebod's feet, literally: to deceive), cannot be written under any of the constituent elements of that idiom, because the above idiom has no any semantic relation with the words *"pust"*, *"xarboze"*, *"pā"*, etc. Therefore, it is obvious that idioms are always displayed as main entries. Some Persian and English dictionaries display idioms under one of the constituent elements of the idiom which helps the user find the idiom quickly. The reason for adopting that method by lexicographers is beyond the scope of this article and we will limit ourselves to this.

Based on the results of the present research, it can be concluded that there are two types of idioms in Persian. Some idioms are inflexible to syntactic constructions and still bear the name of idiom. These idioms will be the main entries in the dictionary as before. But the topic we are discussing is the second type of idioms, which are flexible against syntactic constructions. The authors consider these idioms to be a type of collocation and their representation in dictionaries is similar to collocations, but with two more limitations. Here, the flexible idiom *šāx-e qul-rā šekast-an* (= to break the monster's horn, literally: to do a superhuman task), which will be called a collocation from now on, is represented in the dictionary.

The authors believe that the above expression, which is flexible against syntactic constructions, should not be the main entry. The reason is that its constituent elements were not irreplaceable like idioms and behaved similarly to collocations. The above expression has been recorded in Persian dictionaries as an idiom, but the lexical phrase *"šāx-e qul"* (in *šāx-e qul-rā šekast-an*) has not been inserted in any dictionary as the main entry. This lexical phrase could freely change its place in the idiom construction. One of the reasons, as mentioned, was that it carried all or most of the semantic load of the idiom. This reason is enough to choose "šāx-e qul" as a newly-created lexical phrase as a main entry in the dictionary. The meaning of this new lexical phrase is something close to the concept of "extraordinary work". Since the user cannot freely choose such phrases and use them in his/her text, the definition of these phrases should not be displayed. This is the first limitation in the way these entries are represented. The reason for this is to tell the user that he/she cannot use these phrases alone, except with their collocators. Mandatory use of the collocators is the second limitation in the representation of these phrases. This can be seen in the entry *aback (adv)* in *Longman Dictionary of Contemporary English*, 5th Ed. It can be founded that the use of the adverb *aback* will always be accompanied by the use of the light verb *to take*.

Finally, the way to represent the expression *šāx-e qul-rā šekast-an* is suggested as follows:

šāx-e qul (noun)

~ *rā šekast-an*

In front of the main entry *"šāx-e qul"*, no definition or equivalent like "extraordinary work" has been given because it cannot be used freely in the above meaning. For example, *šāx-ha-y-e qul dastmozd-e bālā-tar-i mi-talab-ad* (= the monster's horns demand a higher salary!) is not possible. In the sub-entry, the sign ~ indicates the presence of the main entry in the sub-entry expression. Using the entry *šāx-e qul* is always accompanied by the use of the collocator *šekast-an*. In fact, by displaying the sub-entry in this way, we allow the user to use the phrase *šāx-e qul-rā šekast-an* in different syntactic constructions such as passivization.

References

Anvari, H. (2019). *Kenayat Sokhan Dictionary* (1st ed.) Tehran: Sokhan.

Atkins, B. T. S. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. New York: OUP.

Croft, W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, W. (2007). Construction grammar. In D. Geeraerts & H. Cuyckens (Eds.), *Handbook of Cognitive Linguistics* (pp. 463-508). Oxford: Oxford University Press.

Dabir-Moghaddan, M. (2020). *Linguistics: An Iranian perspective*. Tehran: Allameh Tabatabai University Press.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.

Goldberg, A. (2003). Constructions: a new theoretical approach to language. *Trends in cognitive sciences*, 7(5), 219-224.
https://doi.org/10.1016/S1364-6613(03)00080-9

Jespersen, O. (1909). *A Modern English Grammar on Historical Principles*. London/Copenhagen: George Allan and Unwin/Munksgaard.

Karimi, S. (1997). Persian complex verbs: idiomatic or compositional? *Lexicology*, 3(2), 273- 318.
https://www.u.arizona.edu/~karimi/publications/Idiomatic%20or%20comp. pdf

Landau, S. I. (2001). *Dictionaries: The Art and Craft of Lexicography*. London: Cambridge University Press.

Mel'čuk, I. (2014). *Semantics: From meaning to text.* John Benjamins Publishing Company.

Miller, J. (1999). Magnasyntax and syntactic analysis. *Revue française de linguistique appliquée*, 4(2), 7-20. https://doi.org/10.3917/rfla.042.0007

Moezzipoor, F. (2018). A Discussion on Persian left-dislocation and its theoretical challenges for Role and Reference Grammar. *Journal of Linguistic Researches in Linguistics*, 2, 45-64. https://doi.org/10.22108/JRL.2019.115388.1308

Najafi, A. (2008). *Farsi Amiyaneh Dictionary* (2nd ed.). Tehran: Niloufar.

Newmeyer, F. (1974). The regularity of idiom behavior. *Lingua*, 34, 327–342. https://doi.org/10.1016/0024-3841(74)90002-3

Perlmutter, D. & P. Postal. (1977). Toward a universal characterization of passivization, *Proceedings of the 3rd annual meeting. Berkeley Linguistic Society*. 394-417. https://doi.org/10.3765/bls.v3i0.2266

Quirk, R., S. Greenbaum, G. Leech, & J. Svartvik. (1985). *A contemporary grammar of the English language.* London: Longman.

Radford, A. (2009). *Analyzing English Sentences, A Minimalist Approach*. Cambridge: Cambridge University Press.

Rundell, M. (2007). *Macmillan English Dictionary for Advanced Learners* (2nd ed.). Oxford: Macmillan Education.

Safavi, K. (2012). *Scattered Writings: The First Book, Semantics.* Tehran: Elmi.

Sanchez, J. & Ott, D. (2020). Dislocations. Language and Linguistics Compass. 14 (9), 1-39. https://doi.org/10.1111/lnc3.12391

Schenk, A. (1995). The syntactic behavior of idioms. In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (eds.), Idioms: Structural and psychological perspectives, 253–271. Hillsdale NJ: Erlbaum Assoc. https://doi.org/10.4324/9781315806501

Sinclair, J. (2004). Trust the Text: Language, Corpus and Discourse. London: Routledge. https://doi.org/10.4324/9780203594070

Sornicola, R. (1988). It-clefts and Wh-clefts: two awkward sentence types. *Linguistics,* 24, 343-379. https://www.jstor.org/stable/4175948

Toma, B. (2018). *The Semantics and Pragmatics of Right Dislocation: Odd thing, that.* Undergraduate Research Thesis. The Ohio State University.

Ward, G. & B. Birner. (1996). A Crosslinguistic Study of Postposing in Discourse. *Language and Speech*, 39, 2-3. https://doi.org/10.1177/002383099603900302