

A CORPUS-BASED DESCRIPTION OF PERSIAN LEXICAL COLLOCATIONS FOR ENGLISH LEARNERS

1. Abrar Hussain Qureshi
2. Mehmood Ul Hassan
3. Shamim Akhter

ABSTRACT:

Collocation is an important level of lexis and its significance in the pedagogy of a language is an established phenomenon. There is also a considerable shift in foreign language teaching from single lexical items to multiword items. As meaning is present in larger stretch of language, the significance of these multi-word items of a language has been increased many times. Anyhow, handling of collocations for the learners of a foreign language is very challenging. The undertaken research is a systematic attempt in this regard to describe Persian language on the level of lexical collocations. The corpus of Persian language, developed by Illinois Institute of Technology, has been used. Sketch Engine, a corpus tool, has been employed to extract lexical collocations from the Persian corpus. The retrieved data of lexical collocations of Persian language is supposed to be useful for Persian bilingual lexicologists, lexicographers, grammarians and in the pedagogy of Persian language.

Key Words: Lexis, collocation, Persian, description, corpus

1. Introduction

The concept of lexis is key to language description. A lexical item has many forms, ranging from single words to compound, idioms, collocation, maxims, etc. مجموعه (word), کتاب سیاه (compound), از این گوش میگیرد, از آن گوش در میکند (idiom), آتش خاموش کردن (collocation), نوبت و ترتیب (maxim) are the best examples of lexical items in Persian language. They are semantic units and carry one potential meaning each. The idea of collocation is a recent one and new in the lexicons of languages. The concept was originated by Firth (1957) and developed by Sinclair (1991). The difference between Firth's and Sinclair's approach was that the former's approach to collocations was not corpus based, while Sinclair's approach to the description of collocation was corpus based. However, it is not easy to define collocation with its complete sphere of functions, even though; many linguists have defined it from various dimensions. In the words of Firth (1957), collocation is the phenomenon of the customary and habitual placement of certain words. While Sinclair (1991) asserts that collocation is the frequent occurrence of two or more words with its arbitrariness. Gairns (1986) asserts that collocation is created because the combination presents a common real life state of affair such as the roaring of the lion rather bellowing of the lion.

¹University of Sahiwal, Pakistan

²Khawaja Fareed UEIT, Rahim yar Khan

³Institute of Southern Punjab Multan

He further explains that the relation between the constituents of a collocation is that of structural or syntagmatic. Nesselhauf (2003) tries to present a clear difference between what is collocation and what may not be called a collocation. He claims that owing to availability of strong measures to determine this difference; it has become easier to refer to the true phenomenon of collocations. He refers it the frequent occurrence of predictable combinations. However, he agrees with the fact that the association of these combination ranges from strong to loose and the advent of corpus and its tools have further made the task easy to determine this strong and loose association.

According to Benson (1986), the study of collocation leads towards two of its types. One is based on the content words, namely, the lexical collocation, while the second can be classified as grammatical collocation, based on the combination of content and grammatical words. Lexical collocations are made by the content words, namely noun, adjectives, adverbs, and verbs for example, “strong tea”, absolutely brave”, etc. The main classes of lexical collocations are adjective noun, verb noun, noun noun, adjective adverb, verb adverb, etc. While the grammatical collocation, known as “colligation” are made of a content word along with a preposition or an infinitive clause. Some of the most frequent grammatical collocations are preposition noun, preposition adjective, noun infinitive, etc.

2.1 The Significance of Collocations

The significance of collocations in language learning has been an established phenomenon. As the lexis is a part of the larger stretches of language utterance and meaning is always present in these larger stretches, the significance of collocations becomes more significant. The meaning of lexical items is always determined if the listener is familiar with the co-text. Moreover, knowing a word means to know the use of the word in the context and without knowing the company of these words, the vocabulary list of a language will be ineffective for language comprehension that is the final end of any language learning program. The importance of these word-combinations in learning a foreign language has been stressed by linguists like Fitikides (2002). As Firth (1957) claims that the speaker comes to know the meaning by its adjacent words.

Every language has a vast majority of synonyms and it is also a fact that no two synonyms have the same meanings. There is a shade of difference between the two synonyms. Though lexicon describes the meanings of the individual lexical items as far as synonyms are concerned but the shade of different meaning can only be determined by collocations. As, “great” and “gigantic” are considered synonyms but both reflect to various degrees of greatness as “gigantic” refers to unbelievable greatness. Another very best example of the synonym is “happy” and “merry” with the heads “Christmas and birthday”. The adjectives happy and merry can be used with Christmas but he the adjective “merry” cannot be used with the “birthday”. Owing to the crucial position of collocation in mastering a language, their significance becomes very established. Proficiency of a language is a matter of learning these fixed multi-word items. They are a useful tool in L1 acquisition and in the second language learning; their role has been significantly established.

2.3 Theoretical Framework:

Lewis's (1997) approach for the description of Persian collocation has been used. He asserts that on the basis of association, collocations can be classified in four categories. According to him, collocations are weak and strong and may be analyzed on the basis of frequency whether low or high. Collocation is strong if the mutual fixedness between the constituents of the collocation is frequent than usual while in weak collocation the mutual fixedness between the constituents of the collocation is not frequent. The strong collocation like جنگ داخلی (civil war) can be considered as tightly linked that function as single lexical item. While اشتباه کردن (make a mistake) is a weak collocations as کردن (to do) occurs with fairly large number of items. Collocations in Persian language may be classified according to the parameters of strong-frequent and strong-infrequent. The previous example جنگ داخلی (civil war) is the best example of strong and frequent collocation in the Persian language. While ورشکست شدن (go bankrupt) is a strong Persian collocation with infrequent co-occurrence of corpus data.

2.4 Previous works on Persian Collocations:

Zareie and Koosha (2002) conducted a research on Persian speaking students and their difficulties in translating English collocations inn Persian language. The data under analysis carefully studied and the errors regarding collocations were retrieved. Five patterns were highlighted. The most problematic area for the learners was the collocations that contained in preposition and the combinations were preposition plus adverbs, adjectives and fixed expressions. The results of the research strengthened the notion that the knowledge of collocation was crucial not only for English but also for Persian as well.

Batani (2010) explores the difference between Persian collocation and Persian idioms. He discusses the nature of Persian collocations and its strong associations between its constituents. He asserts that idioms in the Persian language are more strongly cemented than collocations. He expands the concept by making a comparison with the English collocations and idioms. His research on Persian collocations was not based on large corpus data.

Jabbari and Kavooosi (2017) explore the translation errors that occur due to translating collocation from a foreign language to Persian language. They highlight errors produced by certified translators. For this purpose they analyzed many official translated texts. The data collected was categorized and afterwards statistical analysis of the data was executed.

2.5 Description of Persian Lexical Collocation:

Patterns of lexical collocations are determined on the basis of frequency of corpus data. There have been found many patterns of lexical collocations in Persian language but the most frequent of the patterns are following:

2.5.1 Noun Noun

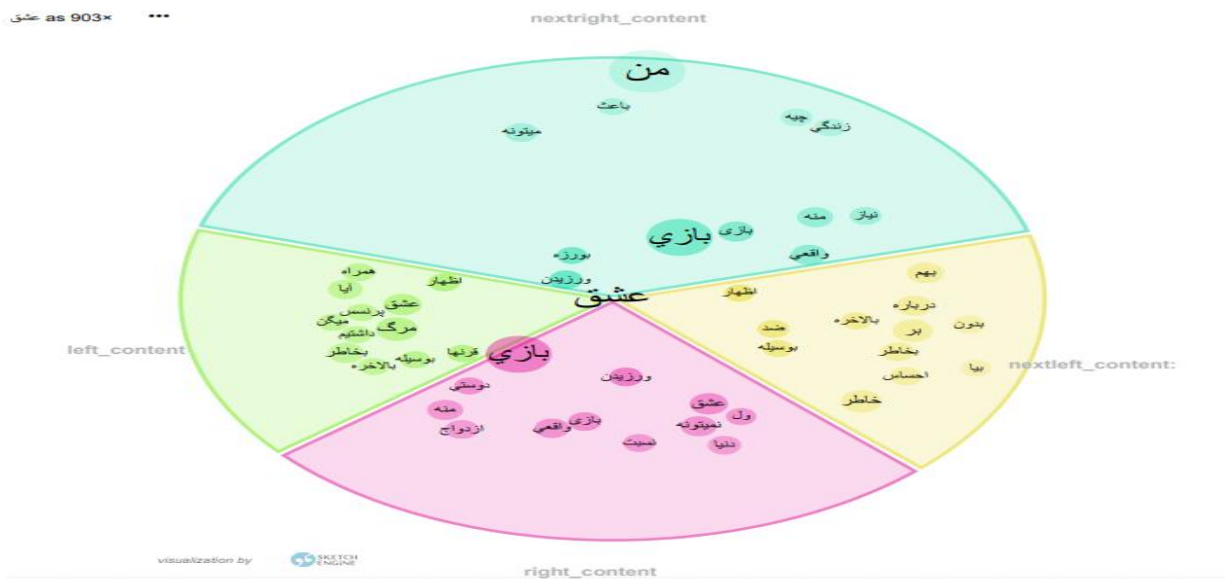
One of the frequent patterns of Persian lexical collocations is noun noun collocation. The following corpus data reflects that the node word "عشق" is frequently accompanied by other nouns on its both sides.

عشق as 903× ...

↔ ☰ ☉ ✕	↔ ☰ ☉ ✕	↔ ☰ ☉ ✕	↔ ☰ ☉ ✕
right_content	left_content	nextright_content	nextleft_content:
بازی ...	اظهار ...	ورزیدن ...	اظهار .
عشق بازی	قرنها ...	بورزه ...	ضد .
ورزیدن ...	عشق ...	بازی ...	بوسیله .
عشق ...	مرگ ...	عشق بازی	بالاخره .
دوستی ...	مرگ بر عشق	بازی ...	بخاطر .
بازی ...	بوسیله ...	واقعی ...	

simple عشق 903 (168.24 per million)

Details	Left context	KWIC	Right context
1	TEP	عشق کمک تو متشکریم. <S></S> به نوشیدنی و غذای مابرتک بده دعا میکنیم همه کسان	<S></S> خداوند. پدر ما، مالک و نجات دهنده از
2	TEP	عشق را داشته باشم. <S></S> اون چیه. <S></S> کجا. <S></S> ایزولد صبر کن. <S></S>	<S></S> و مثل مادرم یک زندگی بدون
3	TEP	عشق به قدر تمندی مرگ است. <S></S> مضحکه. <S></S> علکی باور میکنم چی می	<S></S> همچین روی باز و هات. <S></S>
4	TEP	عشق بورزیم. <S></S> هیچ چیز نمی‌تونه سستش کنه. هیچ چیز نمی‌تونه نابودش کنه.	<S></S> یا تو و من مثل هم بهم
5	TEP	عشق توضیح بدم. <S></S> به نظر میاد میخواد بدون اون زندگی کنه چرا. <S></S>	<S></S> من داشتم سعی میکردم به ترستان در مورد. اهمیت
6	TEP	عشق سپری میکنم. <S></S> عشق را خدا خلق کرده. <S></S> نادیده بگیریش، چنان	<S></S> همشون روز ها روز ها روز ها. بدون
7	TEP	عشق را خدا خلق کرده. <S></S> نادیده بگیریش، چنان رنجی متحمل میشی که نمیتونی	<S></S> بدون عشق سپری میکنم. <S></S>
8	TEP	عشق کور بشه. <S></S> ممکنه یک مردی نتونه بی وفا باشه اگر پشت سر خودش	<S></S> ممکنه باهات صحبت کنم. <S></S> ممکنه، به نظرت یک مرد بوسیله
9	TEP	عشق ورزیدن ما را جرم بدونه. <S></S> سوار شو. <S></S> ترستان، اگر ما اینکار	<S></S> اون هم از ادیمون را به ما داد. <S></S> چرا بهم
10	TEP	عشق ما بود که پادشاهی فرو ریخت. <S></S> ما را به یاد میارن. <S></S> ترستان	<S></S> اگر ما اینکار را نکنیم. <S></S> تا ابد میگویند بخاطر
11	TEP	عشق از هر دو تا شام هم بزرگتر بود. <S></S> صدای ایزولد چهره‌ی من در میان چشمانت	<S></S> نمیتونم اگر زندگی از مرگ بزرگتره. <S></S> اما
12	TEP	عشق بورزیم هیچ چیز نمی‌تونه اونو سست کنه. <S></S> منم، پیتر پارکر. <S></S>	<S></S> یا تو و من مثل هم بهم



The above data shows that noun noun combination of lexical collocation is very frequent in the Persian language. The fact is also noteworthy that nouns occur in lot of combinations of lexical collocations in most of the languages of the world.

2.5.2 Noun Verb

Noun verb is another very frequent word combination in the Persian language. There is one very unique aspect of noun verb combination in the Persian language that a single verb occurs with a variety of nouns. For example the node verb “کردن” occurs with a wide range of nouns.

کردن as 9,277×

left_content	right_content	nextleft_content:	nextright_content
باز	عنوان	مخفی	نهایتگاه
باز کردن	پرونده	مخفی کردن	خطایی ، CRL سعی برای پاک کردن نهایتگاه
مخفی	باز کردن پرونده	باز	سند
مخفی کردن	رشته	باز کردن	باز کردن سند و لاژگان
پاک	نشان دار کردن رشته به عنوان	پاک	پیام
پاک کردن	نشانی	پاک کردن	نشان دار کردن پیام به عنوان
وارد	کردن نشانی وب	دار	رشته
وارد کردن	پرونده	نشان دار کردن	نشان دار کردن رشته به عنوان
مشخص	باز کردن پرونده	وارد	پیامهای
مشخص کردن	جدید	وارد کردن	نشان دار کردن پیامهای برگزیده به عنوان
نشان	نام	مشخص	نام
نشان دار کردن		مشخص کردن	وارد کردن نام

شدن as 2,925x ...

left_content	right_content	nextleft_content:	nextright_content
کشته ...	خنثی ...	کشته ...	خاکستری ...
کشته شدن	خاکستری ...	کشته شدن	دنباله ...
تسلیم ...	دنباله ...	کشته شدن	توسط ...
تسلیم شدن	شدن ...	ناپدید ...	پنجره ...
خلاص ...	دعا ...	ناپدید شدن	ماه ...
خلاص شدن از	قلعه ...	تسلیم ...	آماده ...
باعث ...	پاسخ ...	تسلیم شدن	بر ...
وارد ...	پنجره ...	خلاص شدن از	شدن بر ...
وارد شدن به		برنده ...	ادامه ...
ناپدید ...		برنده شدن	
ناپدید شدن			

simple شدن 2,925 (544.96 per million)

Details	Left context	KWIC	Right context
1	TEP >	شدن دعانمیخونیم. <S><S> دعا واسه همه هست عزیزم پس چر ابرای اونا نمیخونیم.	<S><S> ماتو یادگر فتن آداب مردم دیگر دخالت نمیکنیم. <S><S> برای کسانی که کشته
2	TEP >	شدن در اینجا راه داره. <S><S> من به نیابت از کور نوال. <S><S> پیشنهاد میکنم؛	<S><S> رفت انگیزی بود. <S><S> البته، مردی که مور هلت را شکست میدهد. توانایی پیروز
3	TEP >	شدن تریستان دلاور مارک. من قلعه را فتح میکنم. <S><S> تو میتونی نیرو هات را اونج	<S><S> بچریحه دار میشه. و مقابل فرمانرواهای دیگه دست ننگه میداره. <S><S> باز نجبر
4	TEP >	شدن. <S><S> آکولیان هم به کلی از بین رفته. <S><S> خدای من چطور خائوامون	<S><S> این انتهای دنیاست. <S><S> سیاره کلساک از بین رفته هشت میلیون سکنه ناپدید
5	TEP >	شدن. <S><S> خیلی طول نمیکشه. <S><S> وارد شد. <S><S> ریدیک برگشت.	<S><S> رامون یکی دیگه آوردی. <S><S> پرورش دهنده جالبیه. <S><S> در حال وارد
6	TEP >	شدن شاید استادانه تر باشه. <S><S> در آخر شروع کنی مستقیم برو به هلیون پنج. <S>	<S><S> چونه بیشتر. <S><S> تا وقتی که من با وحشیها مقابله میکنم. <S><S> این نزدیک
7	TEP >	شدن. <S><S> وقت گذاشت. <S><S> حرکت کن. بجنب. <S><S> بجنب. <S>	<S><S> اما الان به نظر میرسه که یک فیوریان استعدادشو داره. <S><S> برای خلاص
8	TEP >	شدن دارم. <S><S> ولی من یکی از رتبه های بالاتر فرماندهها را به دست آوردم. <S>	<S><S>. <S><S> نهو خدا میدونه اینم از روی من بود. <S><S> من خیلی چیزای پاک
9	TEP >	شدن اوضاع. <S><S> و برای تضمین امنیت همگانی. <S><S> بگ وضعت اضطر	<S><S>. <S><S> هرگز شهامت بر آوردنم ندارم. <S><S> برای خاتمه دادن به خراب
10	TEP >	شدن. <S><S> و هنوز تو اینجا نشسته ای و واسه خودت حموم میکنی. <S><S> الیزا	<S><S>. <S><S> بدنام ترین دزدان دریایی دور دنیا. <S><S> همگی بر ضد دشمن ما متحد
11	TEP >	شدن رفته. <S><S> بله جگ، دنیا به او مدن تو احتیاج داره. <S><S> تو به خدمه احد	<S><S> شده. <S><S> مردم را واسه یک دقیقه ها کنید چه اتفاقی افتاده همه چی را به وخیم
12	TEP >	شدن بیگانها از تمام جهان گزارش شده. <S><S> آیا اونها بیگانه هستند. <S><S> ت	<S><S> ز اشیای نورانی در آسمان رسیده. <S><S> با مشاهدات فضاپیما. <S><S> دیده

2.5.3 Noun Adjective Collocation:

Yet another free lexical word combination in the Persian language is noun adjective collocation. This pattern of lexical collocation is quite frequent in the Persian language and constitutes the bulk of Persian lexicon. In the following example, the node "چای" with a number of adjectives, making this combination very dynamic and rich:

چای as 40x ***

right_content	left_content	nextright_content	nextleft_content:
ریونگ 4 ...	مرحم 1 ...	ریونگ 5 ...	مرحم 1 ...
Name 1 ...	درجات 1 ...	زنبیل 1 ...	برگهای 1 ...
Comment 1 ...	برگهای 1 ...	ریونگ 1 ...	پاتگزو 1 ...
ریونگ 1 ...	تفاله 1 ...	پنوشید 1 ...	فتجان 1 ...
بیرونه متوبرید 1 ...	زنجبیل 1 ...	کوهی 1 ...	فتجون 1 ...
قیمتہم 1 ...	میوہ 1 ...	میخورند 1 ...	ویژگیهای 1 ...
مینوشید 1 ...	سیر 1 ...	میوہ 1 ...	کمی 1 ...
فتجانهایی 1 ...	پاتگزو 1 ...	گران 1 ...	مزرعه 1 ...
برنجو 1 ...	فتجون 1 ...	عالی 1 ...	سبز 1 ...
بہخاطر 1 ...	آہستہ 1 ...	امپراتور 1 ...	مزد 1 ...
کوهی 1 ...	ویژگیهای 1 ...	سرد 2 ...	برائون 1 ...
میخورند 1 ...	مخلوط 1 ...	خوشمزہ 1 ...	شاهزادہ 2 ...

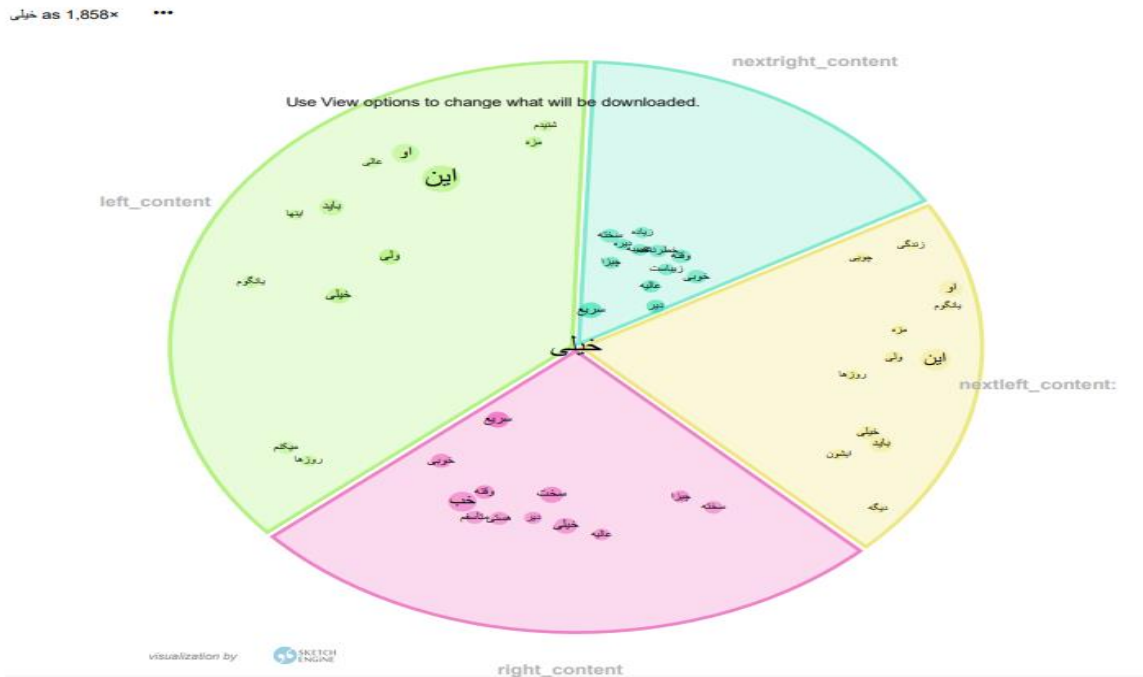
In the following example, the node "شدید" that is adjective, occurs with a wide range of nouns.

word شدید 123 (22.92 per million)

Details	Left context	KWIC	Right context
1 (1) TEP	<</S>>	شدید متمرکز میشی. /><</S>> باعث می شی او چیز سریعتر رخ بده. /><</S>>	<</S>> وقتی روی یک چیزی با احساسات
2 (1) TEP	<</S>>	شدید کم کم احساسش را پیدا می کنید. /><</S>> و اون وقته که با اون یکی	<</S>> یک بر نامه پایه ای طراحی کردیم به نام خود شفا دهی. /><</S>> شما زخمی
3 (1) TEP	<</S>>	شدید با موتور سیکلت داشتیم. /><</S>> یک ز مان من توی دالاس بی خانه مان بود	<</S>> 13 تا 18 سالگی درگیر باند های خیابونی بودم. /><</S>> من یک تصادف
4 (1) TEP	<</S>>	شدید بیرون ساختمان پشتی. /><</S>> به سمت غرب خیابون متوجه	<</S>> بارنز ، تو هستی. /><</S>> میتونید بشینید. /><</S>> زخمی
5 (1) TEP	<</S>>	شدید آره. /><</S>> ولی مهم نیست. /><</S>> مامور توماس بارنز صحبت م	<</S>> قربان. /><</S>> طبق گزارشات از سالامانکار رئیس جمهور. /><</S>> تحت مراقبت
6 (1) TEP	<</S>>	شدید وبم ، یک اشتباه کوچیک در خودشان وجود داشت مثل بچهها. /><</S>> که مذ	<</S>> شحالم دوباره میبینمتون. /><</S>> نمایم داخل. /><</S>> شرمنده معطل
7 (1) TEP	<</S>>	شدید روز. /><</S>> و در سرمای لذت بخش آب. /><</S>> هورتون فیل قصه	<</S>> نظرت میاد که مثل جهنم میسوزه. /><</S>> اون گفت: وقتی که زخم زیاد هم
8 (1) TEP	<</S>>	شدید نیست. /><</S>> جناب سروان ، خواهش میکنم تمومش کن. /><</S>> یا!	<</S>> کنی. /><</S>> نیکول میخوام بهت بگم. /><</S>> آسیب دیدگی من زیاد هم
9 (1) TEP	<</S>>	شدید نه. /><</S>> نه ، اما خیلی دلم می خواد. /><</S>> دکتر ساتو ، مامی تونید	<</S>> رادی نداره. /><</S>> می دونی که خیلی شرم مندم ، /><</S>> که از هم جدا
10 (1) TEP	<</S>>	شدید ما به شما چه گفتیم؟ /><</S>> پیامدهایی هستند که خیانت به پادشاه محسوب	<</S>> تو باید برای سزای گناهت کشته شوی /><</S>> روزی که بانوی در باز

2.5.4 Adverb Adjective Collocation

Yet another free combination of lexical collocation in the Persian language is the adverb adjective collocation. Some Persian adverbs like some Persian nouns occur with a wide range of Persian adjectives for example:



2.5.5 Adverb Verb Collocation:

Adverb verb is another possible word combination of lexical collocation that is present in the Persian language. Though it is not as frequent as the rest of lexical combination yet it constitutes a considerable bulk of Persian lexicon.

اینجا as 2,774x ...

left_content	right_content	nextright_content
بیا 47 7.3 ...	توانید 217 8.62 ...	برگزینید 51 8.19 ...
بیا اینجا	در اینجا می توانید	شود را در اینجا برگزینید
در 692 6.08 ...	چیکار 40 7.26 ...	چیکار 38 8.15 ...
در اینجا می توانید	اینجا چیکار می کنی	اینجا چیکار می کنی
چی 37 5.86 ...	تنظیم 72 7.22 ...	تنظیم 63 7.47 ...
چی چیزی	را در اینجا تنظیم کنید	را در اینجا تنظیم کنید
چیزی 32 5.86 ...	دهید 64 7.09 ...	کشونده 13 7.23 ...
کسی 21 5.55 ...	اینجا را فشار دهید	اینجا کشونده
کسی اینجا	میکنید 37 7.07 ...	آورده 24 7.16 ...
باید 56 5.51 ...	کنید 185 6.98 ...	را به اینجا آورده
باید اینجا	در اینجا تنظیم کنید	اومدی 12 6.97 ...
وقتی 23 5.49 ...	فشار 65 6.73 ...	اینجا اومدی
باتگوم 21 5.48 ...	اینجا را فشار دهید	بریم 17 6.89 ...
باتگوم اینجا	بیرون 36 6.7 ...	از اینجا بریم
نیاید 14 5.45 ...	از اینجا بیرون	آمدید 11 6.88 ...
		آمدید

Another example of adverb verb combination of lexical collocation in the Persian language is the following in which the Persian Adverb occurs with a lot of Persian verbs.



The concept of collocations is a new phenomenon in the Persian language, specially, Persian as a second language. In Persian second language acquisition, the importance of collocations is unchallenged. Handling of the collocations of Persian language contributes towards improving the communicative competence of the language. Bateni (2010) asserts that successful handling of collocations leads towards the mastery of idiomaticity and native-likeness of the target language. He says further that accuracy and fluency of a language is based on these fixed and semi-fixed word combinations of the language and one of the reasons of lexical errors of the foreign language is the lack of proficiency of these phraseological expressions.

3. Conclusion

For Pakistani learners of Persian language, it is very indispensable to have a sound knowledge and mastery over lexical collocations of Persian language if they intend to enhance their proficiency of Persian. A special focus is required to get familiarity with the Persian collocations. With the advent of computers and soft wares, the process of extracting these multi-word combinations has become very simple and easier. Their frequency and strong association can be determined in no time. Rather, it will be better to say, that learning collocation through corpus data has lead toward autonomy of the Persian language learners in the acquisition process.

In the end all the dimensions of Persian collocations may not have been explored. The undertaken research can be taken as a point of departure. It motivates the Persian lexicologists to investigate further word-combinations of the Persian language with the help of larger corpus data, resulting in the emergence of rich and productive linguistic resource of Persian language.

REFERENCES

- Bateni , M.R. (2010) . Collocations and idioms and their translatability, *Journal of Iranian Studies* , 43, 591-597
- Benson, M., Benson, E., & Ilson, R. (1986). *The BBI combinatory dictionary of English: A guide to word combinations*. (pp. i-v). Amsterdam: John Benjamins.
- Firth, J. R. (1951). Modes of meaning. In J. R. Firth (Ed.), *Papers in linguistics* (pp. 115-197). Oxford, UK: Oxford University Press.
- Fitikides, T.J (2002). *Common mistakes in English*. Longman, London
- Jabbari, M.J. and Kavooosi, N. (2017). An Investigation into the Collocations Used in the Translation of Official Documents from Persian into English. *Communication and Linguistics Studies*. 3(2), 15-21/doi: 10.11648/j.cls.20170302.11
- Lewis, M. (1997). *Implementing the lexical approach*. (pp.34-44). England: Language Teaching Publication
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, 24, 223-42.



Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.

Zareie, A & Koosha, M. (2002). Pattern of the Iranian Advanced Learners Problems with English Collocations: A Focus on Lexical Collocations. [Online] Available: http://www.sid.ir/en/VEWSSID/J_pdf (September 2, 2010).