

# Lexical Duplicates: A Special Type of Collocations And How to Teach Them

Hossein Vossoughi Ph.D, hvossoughi78@yahoo.com  
Niloufar Shahmirzadi MA, niloufar\_shahmirzadi@rkiau.ac.ir



## چکیده

هدف از نگارش این مقاله، بررسی لغات هم‌نشین است که به‌رغم کاربردی بودن کمتر مورد استفاده روزمره قرار می‌گیرند. این موضوع را از دو جهت می‌توان بررسی کرد: اول آنکه ممکن است در محیط‌های آموزشی توجه کمتری به تدریس آن‌ها شده باشد و دوم اینکه اهمیت کاربرد آن‌ها در تعاملات روزمره و درک مطالب مورد توجه قرار نگرفته باشد. تقسیم‌بندی این لغات می‌تواند باعث سهولت در یادگیری و بررسی مقابله‌ای آن‌ها شود و فراگیرندگان را به سمت فراگیری هرچه طبیعی‌تر زبان انگلیسی سوق دهد.

**کلیدواژه‌ها:** هم‌نشین‌ها، عبارات‌های ثابت، اصطلاحات، تکرارهای واژگانی

## Abstract

The aim of the present study is to explore a group of special word combinations or bundles of words which come together and conventionally labeled as lexical duplicates. It is also known that the types of combinations could be found both in English and Persian languages despite the different rules that might follow in both of them. To sum up, lexical duplicates should be classified as a separate and special group of them. Moreover, considering the high frequency use of these bundles, it is recommended to put into practice the application of them in daily life so as be able to authenticate flow of speech.

**Key Words:** collocations, fixed expressions, idioms, lexical duplicates

## Introduction

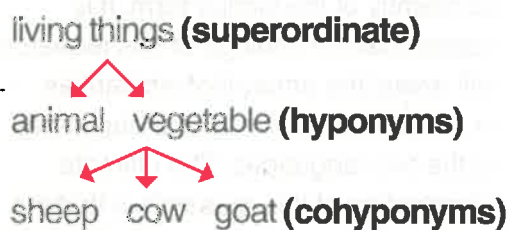
The issues of word have been discussed from different viewpoints. Linguists have formally approached word based on a morphological viewpoint (Akmajian, A. et al. 2001, 1997; Fasold, W.V. & Conner-Linton 2006; Katamba, F. 1993; Yule, G. 1985). In their opinions, words are one of the three main components of language structure,

namely phonology, morphology, and syntax. They maintain that words are made up of one or more meaningful units named morphemes. A word may contain one basic morpheme and a number of other morphemes which are called affixes. The affixes are classified into (i) prefixes if they occur before the base morpheme as "enslave" or (ii) suffixes if they appear after the base part of the

word, such as "closed", "books" and they may be called (iii) superfixes if they are in the form of suprasegmental elements like 'stress' as it can be observed in the nominal form of "record" in contrast with its verbal form "re'cord". (iv) the other type of affix is termed infix because it occurs inside the base or root morpheme.

Morphemes have also been categorized by traditional linguists to free and bound morphemes. A free morpheme can be used as a word or independent lexical item in the language, and it can be assumed that free morpheme is a single morpheme, a word, like *book*, *encyclopedia*, *symptom*, or the base part of the words to which one or more bound morphemes are attached like **exchangeability**, **unsuccessfulness**, **reproductivity**, **representational**.

Semanticists look at words based on the different types of relationships which are held in the different types of semantic networks which exist in the lexicon of a language (Parker, F. & Riley, K. 2005). Most of these relations can be traced in different languages and it seems that they constitute some universal categories. They are distinguished as synonyms, antonyms, hyponyms, superordinates, subordinates, cohyponyms, etc. The examples are in the pairs of words as "wide ~ broad, day ~ night" or in the following diagram:



Another type of study of words is the

description of the type of combinations the words may undergo. This is related to the use and function of words in everyday use of language by the speaker of a language. The **applied linguists** who are interested in looking at how the language is used in its social context, and the **anthropologists**, who are interested in how the culture is reflected in the word combinations or collocations in languages, concentrate on lexical combinations (Schmitt, N. 2002). The applied linguists may have inclination towards cross-cultural linguistics for contrastive analysis or translation activities across languages.

Applied linguists have detected and categorized word combinations into three major groups (Baker, 1992): "fixed expressions", "collocations", and idioms. They usually are referred to as "culture-specific items" (CSIs). The teaching or learning of word combinations in foreign language classrooms is not fashionable or explicit which means learners may get confused when being exposed to them in their task of learning. The question as to how they should be dealt with in the formal education has not yet been even raised in the literature of foreign language teaching. These three combinations are described below with the aim of showing their characteristics and distinctions.

**Collocation:** Baker (1992, p.47) states "collocation would be to think of it in terms of the tendency of certain words to co-occur regularly in a given language." For example in English we usually "*make a mistake*", "*do a research*" and "*pay a visit*" but not "*\*do a mistake*, *\*perform a research*, or *\*do a visit*, etc." Collocates

are a set of words which can co-occur with a word regularly such as the words which have collocated with "dry" in English, such as dry cow, dry voice, dry country, dry humor (Baker, 1992, p.53).

**Fixed expression:** Fixed expressions differ from collocations in that they do not have the flexibility of patterning into other forms or combinations of their constituents. An example of collocation in comparison with a fixed expression may reveal this fact. The collocation "deliver a letter" can be changed into other patterns such as "delivery of a letter", "a letter has been delivered" "having delivered a letter", Baker (1992). they are all acceptable collocations and the meaning of deliver and a letter are transparent in all cases. But a fixed expression is usually "a frozen pattern" of language and the order of the form and of its constituents cannot be changed, such as "ladies and gentlemen", "Merry Christmas", "as a matter of fact", "all the best", "sincerely yours", "every other day", etc.

**Idioms:** They are a type of collocation which are dissimilar from usual collocations and fixed expressions in two consistent ways. First of all, the form and the order of their constituents are not flexible and cannot change, then the meaning of the words which make them up are not transparent or predictable and the number of the constituents of them cannot be changed or reduced. (see Baker, 1992). Some examples are: a deadline (مهلت مقرر), a hot line (تلفن مستقیم بین سران), to pick someone's brain (نظر کسی را جویا شدن), to do the track (به مقصود خود رسیدن), etc. (Sidel & McMordie, 2003).

## The Purpose

The main purpose of this study is to investigate a group of special word combinations or bundles of words which can conventionally be labeled as 'lexical duplicates'. Generally speaking, each combination is mainly made up of two similar lexical counterparts which are duplicate or a pair of identical words namely the second one is the exact mirror reflection of the first one, or it undergoes a slight alternation, or may be attached to the previous one with the intermediation of a type of preposition or a type of conjunction. These special duplicates have not so far been dealt with in any linguistic resources or methodological references of foreign language teaching. Consequently, the familiarization and manipulation of lexical duplicates of English and Persian are of the main concern of this paper.

## Statement of the Problem

Lexical duplicates are surprisingly found both in English and Persian. But the assumption of this research is that the types of their combinations and the kinds of constituents which produce lexical duplicates in one language may not be of the same types of combinations and constituents of the other language or if there are some similar combinations, they may be culturally of the same meaning or accidentally of the similar form. It is hoped that the findings of this research will reveal the amount of similarities or differences of the lexical duplicates of the two languages. The ultimate assumption of this research is that the following relations may be established between these two sets of culture-



specific items as they are predicted by Davies (2004, p.25).

- (a) same form and same meaning (SF-SM)
- (b) different form and different meaning (DF-DM)
- (c) different form and same meaning (DF-SM)
- (d) same form and different meaning (SF-DM)

The procedure of this study is, first of all, to collect, categorize and describe the lexical duplicates in English, then in Persian and finally conduct a kind of contrastive study on them to estimate the degree of their similarities or differences and finally, in the light of the findings, suggest teaching procedures or strategies of learning them by Iranian learners of English as a foreign language.

### Data Collection

The main sources from which these types of lexical bundles or duplicates could be located and extracted were dictionaries, monolingual and bilingual, and the Oxford Dictionary of collocations. The sources for the data collection for collecting the lexical duplicates of the Persian types were Persian monolingual and bilingual dictionaries namely Persian to Persian and Persian to English dictionaries. The researchers had to find the target language equivalents for both groups of lexical duplicates to establish the form-semantic relations for the learners and/or teachers to provide them with the suitable medium for learning or instruction. Some of the English and Persian lexical duplicates are presented here for familiarization, but we leave the classification and analyses of their

patterns for the next section.

English Items	Persian Items
bon bon آبنبات	slowly آرام آرام
chop chop زود زود	snoring خرخر
door to door خانه گرد	hand in hand دست در دست
hush hush خیلی محرمانه	wandering دلی دلی
knick knack خرت و برت	stripped راه راه

### Data Analysis

The data collected from the English language can be generalized into several groups or combination patterns as far as their formal shapes and relations between their two constituents are concerned. In the following section, these groupings or patterns are presented, and described one by one.

- **First pattern:** The first pattern of the English lexical duplicates is the one in which the second constituent is exactly the identical counterpart of the first one and they have occurred side by side without any difference. Seven entities of this type of combination have been found in the data. The following table displays them alphabetically with their Persian translation counterparts.

Table 1. Lexical duplicates of exact constituents

NO.	English (SL)	Persian Equivalents (TL)
1	bon bon	آبنبات
2	boo boo	استباه لبی
3	chop chop	زود زود

- **Second pattern:** In this pattern the initial consonant of the second constituent differs from the initial consonant of the first element of the combination. The other vowels or consonants are the same. Seventeen items of this special form have been

discerned in the whole data. They will be presented in the following table along with their Persian equivalents.

Table 2. Lexical duplicates with different initial consonant of the second one

NO.	English(SL)	Persian Equivalents(TL)
1	boogi woogi	جاز تند
2	booo hoo	زله زار (گریه کردن)
3	fan tan	فان تان (نوعی بازی)

- **Third pattern:** In this combination, the nucleus vowel of the second word differs from that of the first constituent. The typical changes are such as: i → /o/, /u/, /â/, /ou/ or /a/. The number of this type of combinations or lexical duplicates amounts to sixteen items as are shown in the table below:

Table 3. The vowel of the nucleus of the second word differs from that of the first word of the combination

NO.	English(SL)	Persian Equivalents(TL)
1	chit chat	چب / اختلاط
2	criss cross	چب اندر قیچی
3	di do	دسته گل به آب دادن

- **Fourth pattern:** In this pattern, the second constituent is connected to the previous one with the intermediation of the article "a", preposition "to" or the conjunction "and". There may also be a change in the second word like the change in the initial consonants or the change in its nucleus vowel. Ten items of this type of lexical duplicates have been distinguished and extracted from the data. They are presented in the following table.

Table 4. The lexical duplicates with intermediation of "a", "to" or "and" and a probable change in the consonant or vowel of the second element

NO.	English(SL)	Persian Equivalents(TL)
1	bric-a-brac	خرت و برت
2	clock-a-block	مملو / به هم چسبیده
3	day-to-day	روزمره / روزانه

- **Fifth pattern:** These lexical combinations are composed of the couples which are semantically paired, not formally. It is to say that the pairs of words which regularly construct these word bundles are semantically synonyms, antonyms or semantically are not related but they become de-familiarized when they become paired and used in regular combination. It means that two informal or irrelevant words would produce a meaningful and probable idiomatic construction which can be strikingly meaningful and remarkable. These couples are presented in the following table, and the sub-categorization of them are shown in sequential groups and labeled according to their relation types.

Table 5. Semantic coupled pairs of lexical bundles

NO.	English (SL)	Persian Equivalents (TL)
<b>G1 Synonyms</b>		
1	babies and sucklings	کودکان بیگناه
2	bag and baggage	بار و بنه
3	comb and brush	شانه و برس
<b>G2 Antonyms</b>		
4	ancient and modern	در سراسر
5	at home and abroad	در سفر و حضر
6	body and soul	با تمام وجود
<b>G3 De-familiarized Couples</b>		
7	alpha and omega	آغاز و پایان
8	arm and leg	دست و پا
9	art and sciences	علوم و انبیات

## Data Analysis

The total number of the lexical bundles found in the data is 142 items\*, 50 of which are duplicates of the formally similar pairs and the other 92 items are semantically related as synonyms,

antonyms or de-familiarized pairs. The first group can be divided into four subgroups showing a change in dissimilarity of the different initial consonant, dissimilarity of nucleus vowel or both of them, and the combination of two constituents with article, preposition or conjunction. The second subgroup amounts to 92 items, the constituents of the first 9 subgroups are synonymous, the other 21 pairs comprise the antonymous constituents and the last 69 pairs are those which are made up of the two familiar words which produce de-familiarized lexical bundles and their ratio number of each group is given in the following table:

Table 6. The results in percentage terms

Word Bundle	Distribution	Ratio Number
pattern 1	7	4.92%
pattern 2	17	11.97%
pattern 3	16	11.26%
pattern 4	10	7.04%
pattern 5	92	64.78%
total	142	100%

## Persian Special Lexical Bundles or Lexical Duplicates

### • Data Collection

The types of word combinations which are constructed from two, more or less, similar words or lexical counterparts exist in the Persian vocabulary component. The whole data that could be collected for this study amounted to 133 duplicates. The data has already been studied and analyzed and the similar entitles have been generalized into classes to make up special patterns. These patterns are introduced and described one by one in the following section and the related example are given in the relevant tables.

- **First pattern:** In the construction of this pattern, the second constituent of the lexical combination is the exact duplicate or mirror reflection of the first one, such as "آهسته آهسته". The number of this type of lexical duplicate amounts to 77 items which have been extracted from the whole data which have been totally 133 items. They are presented in the following table along with their English translation counterparts.

Table 7. The list of Persian lexical duplicates of similar constituents

English Equivalents (TL)	Persian Combinations (SL)	NO.
slowly slowly	آرام آرام	۱
slowly slowly	آسان آسان	۲
little by little	اندک اندک	۳

- **Second pattern:** The combinations categorized in this group are those in which the second element is connected to the first one by the intermediation of a coordinating conjunction mostly "va"/ (و), whose colloquial alternative is /o/ and rarely "â"/ (آ); or with genitive marker - (e). The total number of them is sixteen in the data. The following table shows them along with their English counterparts. There may be an additional change in the initial consonant or the nucleus vowel of the second constituent, or both, like *khosh o besh*, /kh/ → /b/, and /o/ → /e/

Table 8. The lexical duplicates with coordinating conjunction in between

English Equivalents (TL)	Persian Combination(SL)	NO.
here and there	اینجا و آنجا	۱
this side or that side	این طرف و آن طرف	۲
better or worse	باداباد	۳



- **Third pattern of Persian lexical duplicates:** The combination of this pattern is composed of two duplicate words, the second of which is attached to the first one by the intermediation of a kind of preposition mostly /be/ به and rarely /dar/ در. The number of this type of lexical duplicates added to fifteen items as shown in the following table.

Table 9. Lexical duplicates whose second constituents are objects of preposition

English Equivalents (TL)	Persian Combination(SL)	NO.
first things first	الاهم و فی لاهم	۱
jet lag	آب به آب	۲
side by side	پهلوی به پهلوی	۳

- **Fourth pattern:** This pattern comprises those lexical duplicates that pair each other; however, a derivational or rarely an inflectional suffix is added to the final position of the second one to signal the class or the part of speech of the whole combination. There were about seven of them in the data as shown in the following table.

Table 10. Lexical duplicates with the final affixation

English Equivalents (TL)	Persian Combination(SL)	NO.
in top	بالا بالاها	۱
in rush	هول هولکی	۲
he himself	خود خودش	۳

- **Fifth pattern:** In the combination of the words of this category of lexical duplicates, either a coordinating conjunction plus a change in the nucleus of the second pair or its initial consonant happens or only a change in the vowel or consonant would occurs and the conjunction would be added. The following table some tokens of this type of duplicate pattern.

Table 11. Lexical duplicates with adding coordinate conjunction, and changing of vowel or consonant

English Equivalents (TL)	Persian Combination(SL)	NO.
torn to piece	پاره پوره	۱
routed	ناز و مار	۲
on and off	تق و لوق	۳

The total number of the Persian lexical duplicates collected in the data of this research amounted to 133 items\*, the distribution of each pattern is shown in the following table. The probability of their occurrence is also presented which are demonstrated by their relevant ratio number.

Table 12. The results in percentage terms

Word Bundle	Distribution	Ratio Number
pattern 1	81	52.25%
pattern 2	16	10.32%
pattern 3	21	13.54%
pattern 4	7	4.51%
pattern 5	30	19.35%
total	133	100%

## Findings and Discussions

First of all, by concentrating on the types, forms and semantic aspects of the lexical duplicates, we can infer that they are a class of elements of linguistic compounds which can be categorized as a portion of morphology component and are integrated in the lexicon of the language in the same way as other lexical combinations such as idioms, fixed expressions and collocations are; however, they are a special kind of word bundles which are neither idioms nor fixed expressions but they can characterize both features of them or none of their features. What makes them exclusive and distinctive from the other lexical combinations is the formal relation between their two constituents in which

the second one is the mirror reflection of the first, or the semantic aspect of the second one is the synonym or antonym of the first one.

Secondly, these expressions are amongst the most familiarized and usual daily items that from the one hand carry a heavy load of cultural or social meaning and from the other hand are used in informal, spoken language of daily affairs. In spite of their significance they are not considered in the course of foreign language teaching or learning. They always seem very odd and exotic to the learners of English as a foreign language, because they never appear in the programized and formal language of textbooks except when their presence is highly required in written texts as carrying functional load. For this reason, they are not taken seriously and are disregarded in the program of language teaching. As these items have a great role in authentic conversation in everyday English they should be integrated in the courses of conversation and reading.

Applied linguists have detected and categorized word combinations into three major groups (Baker, 1992): "fixed expressions", "collocations", and idioms. They usually are referred to as "culture-specific items" (CSIs)

### Conclusion and Implications

First of all, lexical duplicates should be categorized as a separate or special group of lexical bundles or collocations other than idioms, fixed expression, and

usual collocations in order to characterize their particularity and their special characteristics in contrast with the other word combinations to be recognized and understood well.

Secondly, the contrastive analysis between the English and Persian lexical duplicates revealed their similarities and differences as enhancing and impeding factors for their teaching and learning. This contrastive analysis and its pedagogical implications require a separate case study which is out of the scope of the present study. It is hoped that this study could have managed to raise the question of lexical duplicates of English and Persian and could have revealed their important role in the teaching and learning of English in state and private schools.

### References

- Akmajan, A., et al. (2001). (5<sup>th</sup> Ed.). *Linguistics: An Introduction to Language and Communication*. Massachusetts: the MIT Press.
- Aryan Pur, K., M. *English- Persian Dictionary, One volume, concise*. Amir Kabir Publications Organization.
- Aryan Pur, K., M. (1984). *The Pocket Persian- English Dictionary, One volume*. Amir Kabir Publications Organization.
- Baker, M. (1992). *In Other Words: A coursebook in translation*. London & New York. Routledge.
- Birjandi, P., Kabiri, Gh., Esmaeel Zare, B., Mahmoodi, Gh. H., Zarei, Sh., Mirhasani, A., Nowruzi, M. (1991). *A Learner's Dictionary of English Idioms*. Ghomes Press.
- Fasold, R.W., & Conner-Linton, J., (2006). (Eds.). *An Introduction to Language and Linguistics*. Cambridge. OUP.
- Haghsheenas, A., M. ( ). *Hezare Dictionary; Two volumes*. Farhang Moaser.
- Katamba, F. (1993). *Morphology*. Hampshire: McMillan Press LTD.
- Moein, M. (1964). *Persian Encyclopedia*. Amir Kabir Publications Organization.
- Parker, F., & Riley, K. (2005). (4<sup>th</sup> Ed.). (Eds.). *Linguistics for Non-Linguists*. Boston: Allyn and Bacon.
- Schmitt, N. (2002). *An Introduction to Applied Linguistics*. Oxford: OUP.
- Seidel, & McMordie. *Translated by Vossoughi* (2003). Tehran, Rahnama Publication.
- Vossoughi, H. (2003). *English-Persian Dictionary of Idioms, (translated)*. Tehran: Rahnama Publication.
- Yule, G. (1985). *The Study of Language*. Cambridge, OUP.
- (2003). (11<sup>th</sup> Ed.) *Merriam-Webster's Collegiate Dictionary*. G. & C. Merriam Company, Springfield, Massachusetts, U.S.A.
- (2004). *Oxford Advanced Learner's Dictionary*. Oxford University Press.

