

The Journal of Teaching Language Skills (JTLS)

4 (4), Winter 2013, Ser. 69/4

ISSN: 2008-8191. pp. 53-78

A Corpus-based Analysis of Collocational Errors in the Iranian EFL Learners' Oral Production

K. Sadeghi *

Assistant Professor, TEFL

Urmia University

email: ksadeghi03@gmail.com

F. Panahifar

M.A., TEFL

Urmia University

email: panahifar.f@gmail.com

Abstract

Collocations are one of the areas generally considered problematic for EFL learners. Iranian learners of English like other EFL learners face various problems in producing oral collocations. An analysis of learners' spoken interlanguage both indicates the scope of the problem and the necessity to spend more time and energy by learners on mastering collocations. The present study specifically focuses on the use of different types of collocations in oral productions of 30 intermediate Iranian EFL learners with the aim of identifying, categorizing and accounting for the inappropriate collocations produced. The corpus analysis revealed that preposition-based collocations in general and verb-preposition collocations in particular are the most problematic types of collocation for Iranian learners of English. The results also indicated that negative transfer from L1 was responsible for a high proportion (56.7%) of collocational errors, while intra-lingual transfer accounted for only 30% of the incorrect collocations. Further findings of the study as well as implications for the teaching and learning of collocations are discussed in the paper.

Keywords: english collocation, negative transfer, Iranian EFL learners, corpus analysis

Received: 05/16/2012

Accepted: 09/16/2012

* Corresponding author

1. Background

Ability to speak a foreign language requires more than mere knowledge of its grammatical and semantic rules. Learners must acquire the knowledge of how native speakers use the language naturally by capitalizing on a wealth of prefabricated forms such as collocations and idioms. Lack of collocational competence often leads “students into grammatical mistakes because they create longer utterances because they do not know the collocations which express precisely what they want to say” (Hill, 2000, p. 49). According to Hill (2000), native speakers are able to speak at the speed they do just for making use of a vast repertoire of ready-made language. Fan (2009, p. 111) attributes production difficulties with collocations to “the idiosyncratic nature of collocational use” and “the fact that collocational use may be markedly different among languages.” She considers the lack of exposure to the target language as the biggest problem resulting in difficulties with the use of collocations.

As far as difficulties in collocational use are concerned, Iranian learners of English are by no means an exception. Any analysis of students’ speech or writing indicates a deficiency in this regard. Despite having sufficient lexical or grammatical knowledge, most Iranian EFL learners seem to experience serious problems with the production of collocational patterns. Such erroneous expressions as *strong rain*, *to take birthday*, *heavy tea*, *to begin a family*, *hard question*, just to name a few, are not due to poor mastery of grammar or lexis. These problems, as Koosha and Jafarpour (2006) have pointed out, stem largely from lack of appropriate collocational knowledge among Iranian EFL learners, the inadequate emphasis given to collocational patterns in their textbooks and the type of instructions they receive. Moreover, such multi-word expressions have not usually been the focus of teaching. The fact is that research professionals in our country and most of the research studies carried out to date (Akbari, 1995; Zarei, 2002; Koosha & Jafarpour, 2006) have investigated the use of collocations in the learners’ written product with only a few touching on their oral performance.

It, therefore, seems essential to identify the problems that EFL learners have in dealing with collocations including their knowledge and use of collocations. Accordingly, the present study tried to push the current status of research on collocations a step forward in attaining this goal. The important role that collocations play in the natural-sounding speech and writing, on the one hand, and problems that Iranian EFL learners have with different types of collocations, on the other hand, highlight the significance of the present study.

More precisely, this study sought answers to the following questions:

1. What are the most problematic types of collocations for Iranian intermediate EFL learners?
2. To what extent is the use of collocations affected by Iranian EFL learners' L1?

2. Review of the Related Literature

2.1 Meaning and significance of collocations

There is no general consensus among linguists on what collocation is, and different definitions have been proposed for the notion of collocation. Nonetheless, most of them are paraphrases of Firth's (1957, p. 183) definition that collocations are "words in habitual company". Cruse (1986), for example, defines collocations as "sequences of lexical items which habitually co-occur, but which are nonetheless fully transparent in the sense that each lexical constituent is also a semantic constituent" (p. 40). Referring to collocations as being easily distinguishable from idioms, he notes that in collocations, there is "a kind of semantic cohesion -- the constituent elements are, to varying degrees, mutually selective" (p. 40) and that in "bound collocations" like *foot the bill*, and *curry favour*, "the constituents do not like to be separated" (p. 41). In his view, although bound collocations show some of the characteristics of idioms, they are nevertheless lexically complex.

Richards and Schmidt (2002, p. 87) define collocation as "the way in which words are used together regularly." Based on this definition, collocation refers to the restrictions on how words can be used together; for example, which verbs and nouns go together, or which adjectives are used with particular nouns. For example, in English the verb *do* collocates with *damage*, *duty*, and *wrong*, but not with *trouble*, *noise*, and *excuse*. Similarly, *high* collocates with *probability* but not with *chance*. We say *high probability* but a *good chance*.

Lewis (2000) defines collocation as "the way in which words co-occur in natural text in statistically significant ways" (p. 132). For Nattinger and DeCarrico (1992), collocations are defined as "strings of specific lexical items that co-occur with a mutual expectancy greater than chance, such as *rancid butter* and *curry favour*" (p. 36). For James (1998), collocations are "the other words any particular word normally keeps company with" (p. 152).

McCarthy (1990) believes that collocation is "an important organizing principle in the vocabulary of any language" (p. 12). For him, collocational knowledge includes part of native speakers' competence, and can pose

problems for EFL learners in cases where collocability is language-specific and is not solely determined by universal semantic restrictions. Pointing to multi-word expressions as an essential component of fluent linguistic production and also a key factor in successful language learning, Hyland (2008) emphasizes the importance of collocations in this way:

An important component of fluent linguistic production is control of the multi-word expressions referred to as clusters, chunks or bundles. These are extended collocations which appear more frequently than expected by chance, helping to shape meanings in specific contexts and contributing to our sense of coherence in a text. (p. 4).

According to Hill (2000), besides being familiar with the concept of communicative competence, the notion of collocational competence is necessary to be added to our thinking. As he puts it, “within the mental lexicon, collocation is the most powerful force in the creation and comprehension of all naturally-occurring text” (p. 49), and it includes one of the most significant areas of idiomatic language. Viewed in this way, collocation is supposed to occupy a central place in the applied linguistics research. However, it was only recently with the advent of corpus linguistics that research into vocabulary in general and collocation in particular has blossomed (Harmer, 2001; Schmitt, 2002; Boers, et al., 2006).

2.2 Collocations and language transfer

Interlingual transfer is considered a significant source for most L2 learner problems (see Brown, 2000; Erdogan, 2005; Mahmoud, 2005). Selinker (1972; as cited in Shahheidari-pour, 2000) proposed negative language transfer as one of the five central processes existing in a latent psychological structure which could be activated in order to learn another language after the end of the critical period for language acquisition.

Errors made due to negative transfer from the first language may occur at all levels of linguistic analysis such as phonology, syntax, lexis and grammar (Oldin, 1989). No doubt, one of the areas susceptible to interlingual transfer is collocation, especially where the first language does not correspond with the target language in terms of collocational patterns. As Sadeghi (2009) rightly points out, a significant number of syntactic and semantic errors made by EFL learners (and sometimes their teachers) may result from a discrepancy between collocational patterns in the L1 and the target language they are struggling to learn. For example, in English they 'smoke cigarette', but in Persian they '*draw cigarette', in English they have

'false teeth', but in Persian they have '*artificial teeth'. Hence, Persian learners of English who are not competent enough in English language may run into the pitfall of negative transfer from their first language and produce incorrect word combinations.

Although study of interlingual transfer as a significant source of errors dates back to 1950s as behaviorist learning theory emerged, it still remains one of the important issues in second language research. In a study on assessment of collocational knowledge of 200 third-year and fourth-year students majoring in English at Yarmouk University, Hussein (1990) found that EFL learners' collocational errors could be traceable to several factors including their unfamiliarity with the structure of collocations, overgeneralization and negative transfer from the first language.

Analyzing collocation errors made by Arab learners of English, Mahmoud (2005) also found negative transfer from the first language as a major source of collocational errors in the writings of Arab EFL learners. In his study, a total of 420 collocations (grammatical and lexical) were extracted from 42 essays written by Arabic-speaking university students majoring in English, of which 269 (64%) were incorrect. More surprisingly, about 61 percent of the incorrect word combinations were made due to negative transfer from Arabic.

In a comparative corpus-based study, Fan (2009) attempted to investigate collocational use by ESL learners compared to that by native students. The data for this study came from written productions of 60 Hong Kong students and 60 native students of English. The corpora analysis revealed that the first group's collocational use was adversely affected by their L1.

As a part of their study to investigate the effect of Iranian EFL learners' L1 on the collocational knowledge of prepositions, Koosha and Jafarpour (2006) used a translation task comprising sixty fill-in-the-blank items on the collocation of prepositions. A total of 4365 errors were extracted, of which 68% was due to inter-lingual errors and 31% stemmed from intra-lingual transfer. Based on the findings, they concluded that first language interference was a significant cause of error in the production of prepositional collocations.

First language interference in the production of collocations was also corroborated in the work of Sadeghi (2009). In a comparative study with 76 participants to investigate the effect of collocational differences between the two languages (i.e., Persian and English) on the participants' collocational errors, he found that a large proportion of collocational problems (about

85%) of all the three groups of different proficiency levels (i.e. low, mid, and high) was due to L1 interference.

2.3 EFL learners' problems with collocations

In his comparative study on collocational clashes to investigate the difficulty levels of different types of collocations Iranian EFL learners and translators face, Nowruzi Khiabani (2000) categorized Adjective-Noun and Noun-Noun collocations into six types irrespective of their syntactic patterns. The first type includes collocations in which the component lexical items are used in their primary sense, and there are lexical equivalents for those primary meanings in the target language, e.g. *military intervention*, *experimental stage*, and *racial discrimination*. For him, the EFL learners' reliance on their first language in order to produce this type of collocations will not cause any problem. The production of collocations becomes problematic when at least one of the word partners appears in its non-primary sense that constitutes the second type of collocations such as *dead market*, *heavy accent*, *easy money*, and *dead silence*. The third type includes word groups such as *finishing blow*, *liberal arts* in which none of the lexical items are used in the primary sense. So this type, according to Nowruzi Khiabani, produces more problems to Iranian learners of English than the previous type for having no clue for elicitation of the target language elements. The fourth type consists of collocations whose meanings are expressed by only one word in Persian language e.g. *military police*, *fringe benefit*, *high tide*, *foreign exchange*. Included in fifth type are collocations such as *jury*, *accomplice*, *appointment*, *agenda*, in which the Persian EFL learners have to collapse several forms to one form in the target language. The last type of collocations comprises word combinations whose correct use requires background information on the part of the learners. Some of these collocations, which according to Nowruzi Khiabani (2000) cannot be categorized under proper names or idiomatic expressions, are: *Achilles heel* (weak point), *fifth columnist* (spy), *Number Ten* (British government) and *Sword of Damocles* (imminent danger).

Akbari (1995) conducted a study in order to test the hypothesis that language-specific collocations are among the major sources of errors in the written production of foreign language learners. His study was conducted with 145 intermediate English majors who were required to write essays during a whole semester. A total of 725 compositions were analyzed for the purpose of this study. The error analysis of the data showed that 47.8% of the errors were grammatical and 52.2% were lexical with collocational errors constituting 10.5% of the total. Among the incorrect collocational

patterns observed in the research (i.e. Adjective-Noun; Noun-Noun; and Verb-Noun), the first one with 6.1% was the most mistaken type of collocational patterns.

Zarei (2002) classified English collocational patterns into ten categories among which the prepositional collocations were the most problematic patterns, while “adverb + adjective” and “fixed expressions” ranked the least problematic. Zarei concluded that knowledge of collocations was an essential part of achieving native-like competence in English.

Based on their study with 200 participants, Koosha and Jafarpour concluded that the majority of Iranian EFL learners had good knowledge of English grammar and vocabulary; however, they seemed to have serious problems with the production of collocational patterns, especially collocations of prepositions. They further found that prepositional collocations did not exert the same degree of difficulty on different levels of proficiency levels among Iranian EFL learners.

In his study to investigate the collocational errors in EFL college learners' writing in Taiwan, Li (2005) found that different types of collocations posed different degrees of difficulty for Taiwanese EFL learners. The data for his study came from 38 sophomores. A total of 188 collocational errors (121 grammatical and 67 lexical) were found using error analysis. The results showed that (V + N) and (V + Prep + O/ V + O + Prep + O) errors occurred most frequently in the participants' writing and (Adj + to infinitive) errors were the least occurring ones. However, the participants' ideas in a questionnaire yielded a different result: the participants considered the patterns (Prep+ N) and (Adj + Prep) the most difficult patterns and (V + O + to be + C/ V + O + C/ V + C) and (V + to Inf/ V + O + to Inf) the easiest ones (where V stands for Verb, N for Noun, O for Object, C for Clause, Prep for Preposition, and Adj stands for Adjective). Li found the ignorance of rule restrictions as the major source of collocational errors in the written productions of Taiwanese' EFL learners.

Studying the relationship between collocations and coherence in the writings of Chinese non-English and English majors, Ying (2009) identified the characteristics of collocations in their compositions. He found that both groups made far greater mistakes in lexical collocations as opposed to grammatical collocations. However, the difference in the wrong use of collocations in their writings appeared in terms of the types of mistakes made by each group. Non-English majors produced more “preposition-noun” type of collocational errors while English majors made more “verb-noun” type of mistakes.

As the above review shows, most of the previous research into collocations has investigated the use of collocations in the learners' written product with only a few touching on their oral performance. Accordingly, the current study aimed at investigating EFL learners' oral use of collocations to examine the role first language plays in oral use of collocations and to identify the most challenging types of collocation.

3. Method

3.1 Participants

The study was conducted with 30 intermediate EFL learners selected from 45 students using an institutional version of PBT TOEFL. The participants included 12 undergraduate students studying English Translation at Payame Nour University of Bonab, and 18 more EFL learners enrolled in Kish Foreign Language Learning Institute in Tehran. The researchers aimed to include all the participants from the same context (that is, from Payame Nour University of Bonab); however, the number of participants in that institution was not enough for our purposes, and we had to recruit similar participants (in terms of language proficiency, the variable which could contribute to differential performance in the production of collocations) from somewhere else. The ease of access to participants at Kish Institute and the similarity of their general language proficiency were the main motivations behind recruiting them in the study, although better generalizability could be achieved were we able to find more homogeneous participants with similar study levels and age.

First language (L1) of 18 participants was Persian and that of 12 participants was Azeri. They were between 17 and 34 years of age. As regards the gender, 28 of the participants were female and 2 were male and that is why we were unable to regard gender as a moderator variable in this study. The distribution of the participants in terms of their gender and L1 is presented in Table 1.

Table 1. Characteristics of the participants

Group	N	Average Score on TOEFL	L1		Gender	
			Azeri	Persian	Male	Female
University students	12	38	12	0	2	10
Institute students	18	40	0	18	0	18
Total	30		12	18	2	28

3.2 Materials

3.2.1 Learners' corpus

The data for this study came from a corpus of 30 spoken productions of EFL learners using semi-structured interviews on a range of topics (Childhood, Language, Shopping, Money, and Employment). The topics were selected based on IELTS' (International English Language Testing System) recommended topics and were meant to be of interest to the participants.

3.2.3 Dictionary of collocations and BNC

In order to decide on the acceptability of the learner-made collocations, and because no native speakers were available who could help with such decisions, *Oxford Collocations Dictionary for Students of English* which gives access to 250,000 word combinations and 75,000 examples of how these collocations are used, as well as the *British National Corpus* (available at <http://www.natcorp.ox.ac.uk>) were used to provide information on the words that headwords could combine with. The *British National Corpus* comprises 100 million words of written and spoken language; the written part includes 90 million words from eight genres and the spoken part consists of 10 million words from four social-class groupings. BNC is one of the most important research tools currently used for corpus-based study of English.

3.2.4 TOEFL

An institutional version of paper-based TOEFL (Sharpe, 2009) was used to determine the proficiency level of the participants. The TOEFL test was judged to be a suitable test for students majoring in English Translation at Payame Nour University of Bonab as they were expected to enjoy an upper-intermediate proficiency level. Initially, all the participants were expected to be recruited from Bonab research site; however, due to lack of enough participants, a decision was made to include language learners of similar proficiency level from a different context (Kish language institute, for the reasons mentioned above). The fact that learners in this second research site were advanced candidates made TOEFL appropriate for them as well. The test was adapted to cover only Grammar and Reading as the testing conditions and the fact that participants came from two institutions made it next to impossible to administer a listening test. The TOEFL used consisted of 60-item multiple choice items with 25 grammar items, and 4 reading passages followed by 35 comprehension questions.

3.3 Procedure

A total of 45 participants including 20 BA students from Payame Nour University of Bonab and 25 EFL learners enrolled in one of Tehran's

branches of Kish Language Learning Institute were invited to take part in the study. At first, in order to ascertain that the participants were on a par as far as language proficiency at the desired level (intermediate level) was concerned, the TOEFL test (as described above) was administered to determine their level of proficiency. The test had 60 items and took test-takers 50 minutes to complete. A total of thirty participants whose scores were between the mean \pm 1 SD were chosen and invited to attend an interview. The participants were individually interviewed by the researchers and were asked to speak about a range of topics (Childhood, Language, Shopping, Money, and Employment). Each participant was asked more than 30 questions regarding these topics. The interviews followed the style used in IELTS speaking module and lasted between 10 and 15 minutes each. With the participants' permission, the interviews were recorded using an MP3 recorder.

3.4 Data analysis

In order to ensure score reliability of the participants' spoken productions, the participants' spoken data were analytically rated by two raters using Weir's (2005) speaking rating criteria. The raters were one of the researchers and a lecturer at Payame Nour University of Bonab. Because of having no access to native speakers, *Oxford Collocations Dictionary for Students of English* which gives access to 250,000 word combinations and 75,000 examples of how these collocations are used, as well as the *British National Corpus* which comprises 100 million words of written and spoken language were used to guide the raters on the words that headwords could combine with. All of the collocations used in the corpus could be found in the collocation dictionary and BNC. The inter-rater reliability was estimated to be $r = 0.913$.

In the next stage, the interviews were first transcribed and all learner-made word combinations were manually extracted from the corpus. Having extracted all word combinations, the researchers separated collocations from free word combinations. For example, regarding verb-noun collocations, if the dictionary (*Oxford Collocations Dictionary for Students of English*) gave no indication of a possible restriction on the verb or sense of the verb in question, i.e. if no specific noun or nouns were given in the definition and there was no other indication of restriction such as individual examples separated by slashes (e.g. *take a vote/poll/survey*), the verb was considered as having unrestricted sense and the combination was classified as free. On the other hand, if there were clear indications that the use of the verb in a particular sense was restricted to a few nouns, it was considered as having a

restricted sense and the combination was classified as a collocation. The same procedure was applied to other types of word combinations such as adjective-noun, adjective-preposition, noun-preposition, verb-preposition, adverb-adjective and verb-adverb.

For the purpose of this study, collocations were defined as two words belonging to different categories (to exclude “noun-noun” collocations and binomials where the two words are from the same category) and are connected by a conjunction or a preposition such as “in” or “by” (e.g. *life and death, here and there, hand in hand, more or less, sooner or later*).

In the second step of the analysis, the extracted collocations were classified into seven categories. In the third step of the analysis, a number of tools were employed in order to judge the acceptability of the collocations produced by the learners. Collocations were judged correct if they were found in the same form in the *Oxford Collocations Dictionary for Students of English*, or in the *British National Corpus*. In the final part of the analysis, based on the collocational errors made by the participants in the spoken data, the most problematic types of collocations for Iranian EFL learners were identified and tabulated. The possible influence of the learners' L1 on the production of English collocations was also assessed on the basis of their possible Persian/Azeri equivalents. The role of the learners' first language was only investigated as regards wrong collocations, thus ignoring the possible positive influence of Persian and Azeri on learner-made English collocations.

4. Results

A content-analysis of the spoken corpus was undertaken in order to describe the possible collocational errors in the oral productions of Iranian learners of English and to answer the research questions pose above. To this end, all word combinations were manually extracted from the corpus and a total of 790 collocations were identified in the corpus.

The collocations produced by the participants yielded seven distinct patterns: 1) verb-preposition, 2) noun-preposition, 3) adjective-preposition, 4) verb-noun, 5) adjective-noun, 6) adverb-adjective and 7) verb-adverb. Based on this categorization, the first three patterns constitute “grammatical collocations” while the remaining four are “lexical collocations”. Table 2 displays the syntactic categorization of the collocations extracted from the EFL learners' spoken corpus.

Table 2. Syntactic patterns of collocations

	Type of Collocation	N	Examples from the Corpus
1	Verb+ Preposition	137	Depend on, argue with, bring about, deal
2	Verb+ Noun	172	with, struggle with
3	Noun+ Preposition	108	Pay attention, raise price, commit crime,
4	Adjective+ Noun	190	get information
5	Adjective+ Preposition	51	Freedom in, saying about, education in,
6	Adverb+ Adjective	92	change in
7	Verb+ Adverb	40	Cute child, clear conscience, high expectation, outstanding role Boring for, good at, ashamed of, crazy about Really ready, completely different, very naughty Change a lot, study hard, improve gradually

As the Table displays, a total of 790 collocations were extracted from the corpus of 30 oral productions, with “adjective-noun” and “verb-adverb” collocations comprising the most and the least frequently occurring types of collocations, respectively. It also shows that most of the collocations produced by the participants were lexical (62.53%), while grammatical collocations (i.e. those including prepositions, accounted for 37.47% of the collocations).

4.1 General categories of collocational errors

The corpus analysis revealed 150 cases of collocational errors. Out of seven types of collocations, “verb-preposition” and “verb-noun” categories with 41 collocational errors each accounted for the most mistakes. However, as far as the ratio of errors to the total number of collocations produced in each particular category was concerned, the major collocational error categories were as follows: (1) verb-preposition, (2) verb-adverb, (3) verb-noun, (4) adjective-preposition, (5) adjective-noun, (6) noun-preposition and (7) adverb-adjective. The figures expressed in terms of percentages in Table 3 indicate the proportion of errors to the total number of collocations in the relevant category.

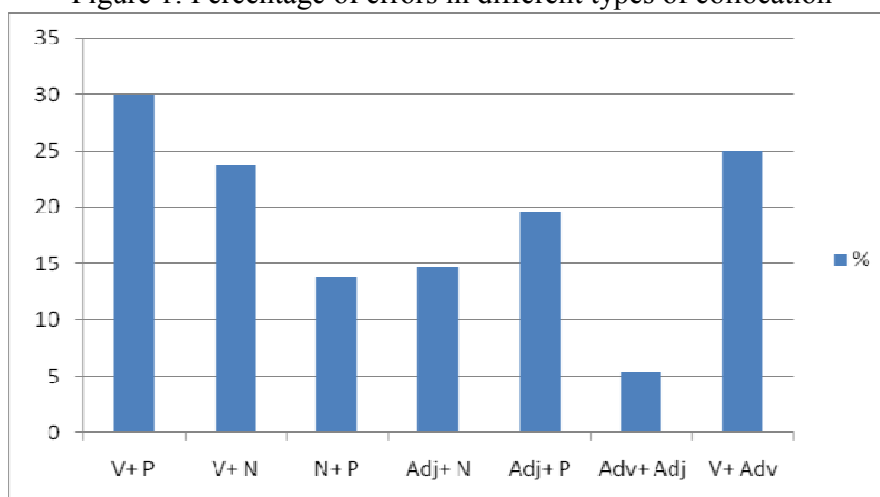
Table 3. Frequency and percentage of collocational errors

Type of Collocation		N	Frequency of Errors	%
1	Verb+ Preposition	137	41	29.9
2	Verb+ Noun	172	41	23.8
3	Noun+ Preposition	108	15	13.8
4	Adjective+ Noun	190	28	14.7
5	Adjective+ Preposition	51	10	19.6
6	Adverb+ Adjective	92	5	5.4
7	Verb+ Adverb	40	10	25.0
Total		790	150	18.9

N= Total number of collocations per category

As the data in Table 3 (and Figure 1 below) show, a total of 150 collocations, out of 790 collocations produced by the participants, were judged unacceptable in English; 84 (56%) of these were lexical and 66 (44%) were grammatical. Of 84 incorrect lexical collocations, 41 (49%) were verb-noun collocations, 28 (33%) were adjective-noun collocations, 10 (12%) were verb-adverb and only 5 (6%) were adverb-adjective collocations. In incorrect grammatical collocations (i.e. verb-preposition, noun-preposition and adjective-preposition collocations), the errors were due to omission, addition and selection of a wrong preposition which accounted for 11%, 28%, and 55% of the errors, respectively.

Figure 1: Percentage of errors in different types of collocation



Following Nesselhauf (2003), all of the incorrect collocations identified in the corpus were classified in terms of types of mistakes. Table 4 represents the seven types of mistakes which were identified for the 150 incorrect collocations categorized also in terms of L1 background of the learners.

Table 4. Types of collocational errors

Type of error		Examples	Occurrence		
			Persian	Azeri	Total
1	Verb Wrong choice of verb	*Technology go up (develop), *See dreams (have), *Gather money (collect), *Reach ones' wish (get)	23	16	39
2	Noun Wrong choice of noun	*Low right (salary), change neighborhood (move house)	1	1	2
3	Adjective Wrong choice of adjectives	*Favorite wish (greatest), *Huge wind (strong), *Good institution (powerful)	15	11	26
4	Adverb Wrong choice of Adverb with verbs or adjectives	*Practice a lot (hard), *Speak a language right (fluently), *Use it very good (properly)	9	6	15
5	Preposition Preposition is missing or existent but wrong	*Order to (#), *Hurt to (#), *Nag about (at), *Listen music (to), *Play by computers (with)	30	34	64
6	Usage Combination does not exist and cannot be corrected by exchanging single elements	*Jump from this job to another job, *limit child, *become a doctor for women (being a gynecologist)	1	3	4
7	Structure Wrong syntactic structure	*Have a argue with (have an argument), *have not any fail (have not failed)	2	0	2

As the data in Table 4 above display, the participants experienced the most problems choosing the correct preposition. Additionally, the wrong choice of verb and adjective accounted for the second and third most frequently occurring types of mistakes, respectively. However, the choice of noun posed little difficulty for them; neither did the choice of adverb (especially in the adverb-adjective combinations) trouble the participants, mainly for having the intensifier “very” at their disposal.

4.2 Collocational errors across learners with different L1

The analysis of the learner-made collocations revealed the Persian group made more correct use of English collocations. The results of this analysis are summarized in Table 5.

Table 5. Collocational errors made by learners with different L1

L1	Persian			Azeri		
	Frequency	N	Errors%	Frequency	N	Errors %
V + P	83	22	26.5	54	19	35.1
V + N	102	24	23.5	70	17	24.2
N + P	49	4	8.1	59	11	18.6
Adj + N	108	15	13.8	82	13	15.8
Adj + P	26	5	19.2	25	5	20.0
Adv + Adj	56	2	3.5	36	3	8.3
V + Adv	28	7	25.0	12	3	24.2
Total	452	79	17.4	338	71	21.0

N = Total number of collocational errors per category

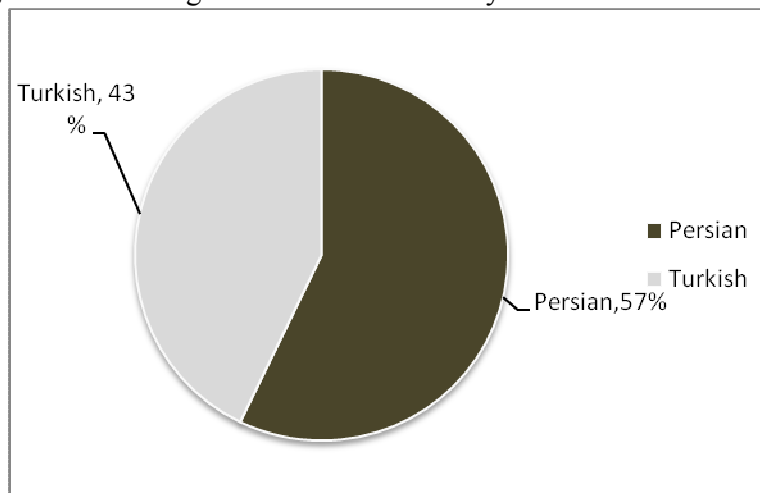
F = Frequency of collocations per category

As the results show, the analysis of the oral production of Persian and Azeri participants recorded 17.4 and 21 percent of collocational errors for each group, respectively. The Persian group's ability to form native-like collocations was most apparent regarding prepositional collocations. As the Table displays, the percentage of errors in verb-preposition and noun-preposition categories is 26.5% and 8.1% for Persian group, while Azeri participants committed 35.1% and 18.6% errors, respectively in categories in question. A chi-square test was run to investigate the significance of these differences per category. The results of chi-square test ($X^2= 6.45$, 1df, $p = 0.002$) demonstrated a significant difference between two groups as to collocational errors made in noun-preposition category. However, the results of chi-square test showed that the difference between two groups' performance as regards other types of collocations was not significant.

As far as the number of collocations produced by each group was concerned, the Persian group with 57% oral use of collocations ranked first, while the Azeri (Turkish) participants produced 43% of collocations in their oral productions (Figure 2). A chi-square was also run to measure significance of these differences in each category. The chi-square results ($X^2= 10.77$, 1df, $p = 0.000$) indicated a significant difference between two groups in terms of frequency of collocations in noun-preposition category, while the difference between two groups' performance as regards other types of collocations was not statistically significant.

Such a difference in their performance can be traced back to the method and kind of English training they receive. The Azeri participants were Payame Noor (distance education) university students majoring in English Translation: hence, they significantly dealt with specialized courses and they did not attend regular classes where general English could be taught or practiced, but the Persian participants studied English at a language institute where the opportunity to attend to and learn collocations increases to a great extent. Furthermore, like English, Persian language counts 'prepositions' as a distinct word class which makes them more 'noticeable' when compared to English equivalents, contributing to their learning; however, Azeri language used an inflectional system for prepositions making prepositions parts of the words to which they are attached rather than as distinct words.

Figure 2: Percentage of collocational use by learners with different L1



4.3 Sources of errors in learner-made collocations

Richards (1974) classifies the errors the language learners make into two distinctive sources: (1) interlingual errors (as the result of the learner’s first language interference) and (2) intralingual errors (which are caused by overgeneralization of L2 rules). According to James (1998), there are four causes of errors: interlingual errors, intralingual errors, communication strategy-based errors (which occurs when learners use some near-equivalent L2 items to replace the target item during communication in L2), and induced errors (the errors caused by the way a teacher gives definitions, examples, explanations or arranges practice opportunities). However, we have used Richards’ categorization as well as a third category which we have added ourselves, namely incorrect collocations due to *unfamiliarity with collocation structure*, to delve into collocational errors found in the corpus in an attempt to identify their sources.

4.3.1 Incorrect collocations due to interlingual transfer

Table 6 shows the frequencies and percentage of incorrect collocations traceable to interference from the first language.

Table 6. Collocational errors traceable to interlingual transfer from L1

Type of Collocation		Persian Speakers			Azeri Speakers					ΣT	ΣL1	Σ%	
		T	L1	%	T	L1							
						P	A	PA	T	%			
1	Verb+ Preposition	22	14	63.6	19	4	5	4	13	68.4	41	27	65.8
2	Verb+ Noun	24	15	62.5	17	0	0	8	8	47.0	41	23	56.0
3	Noun+ Preposition	4	1	25.0	11	0	0	1	1	9.0	15	2	13.3
4	Adjective+ Noun	15	10	66.6	13	4	0	6	10	76.9	28	20	71.4
5	Adjective+ Preposition	5	0	0.0	5	3	0	0	3	60.0	10	3	30.0
6	Adverb+ Adjective	2	0	0.0	3	0	0	0	0	0.0	5	0	0.0
7	Verb+ Adverb	7	7	100	3	1	0	2	3	100	10	10	100
Total		79	47	59.5	71	12	5	21	38	53.5	150	85	56.7

T= Total number of errors per category

L1= Number of errors due to interlingual transfer

P= Persian language A= Azeri language PA= Persian and Azeri languages

As the data in Table 6 display, 56.7 percent of the collocational errors are due to negative transfer, suggesting that learners’ L1 significantly affects the production of collocations. The percentage of negative transfer is highest

(100%) in verb-adverb category and lowest in adverb-adjective category (0.0%). The following examples provide some evidence of incorrect collocations affected by the learners' L1:

Table 7. Incorrect collocations traceable to the learners' L1

Type	Learner-made collocation	Target collocation
V+ P	Take me school	Take me to school
V+P	Regret about	Regret #
N+ P	Opportunity from	Opportunity by
V+ Adv	Speak a language right	Speak a language fluently
Adj+ N	Low right	Low salary
V+ N	See dreams	Have dreams

As Table 6 shows, out of 38 interlingual collocational errors made by the Azeri learners, 13.1 percent is due to Azeri language transfer, 31.6% results from Persian and 55.3 percent can be traced back to both of them. Based on the data, it seems that Azeri learners appealed to the Persian language more than to the Azeri language. It is not surprising since Persian is the language of media and it is widely used by teachers and professors at schools and universities for education. Moreover, a large number of Azeri children are taught Persian (by their parents) even prior to acquiring their L1. However, to draw firm conclusions in this regard requires further studies. The following examples provide some evidence of such negative transfer.

Table 8. Collocational errors due to transfer from Azeri

Type	Learner-made collocation	Target collocation
V+P	Look our pockets	Look at our pockets
V+P	Jump from this job to another job	Change jobs
V+P	Think for themselves	Think about themselves

Table 9. Collocational errors due to transfer from Persian

Type	Learner-made collocation	Target collocation
V+ P	Regret about	Regret #
V+ P	Enter to	Enter #
Adj+ P	Bored from	Bored with
V+ Adv	Fight verbally to	speak harshly to

Table 10. Collocational errors attributable to both Persian and Azeri

Type	Learner-made collocation	Target collocation
V+ P	Moved Ajabshir	Moved to Ajabshir
V+ P	Listen music	Listen to music
V+ Adv	Grow intellectually	Develop intellectually
Adj+ N	Busy child	Naughty child

As mentioned earlier, one aim of this study was to investigate the role of the participants' first language in the production of wrong collocations. However, the positive influence of Persian on learner-made collocations should not be undervalued. Such a positive influence is self-evident in examples like "high expectation", "entrance examination", "bitter experience" and "pass the problems", just to name a few.

4.3.2 Incorrect collocations due to intralingual transfer

As Hussein (1990) points out, a major strategy generally adopted by EFL learners is the reduction of the target language to a simple system which is realized through generalizations. For him, "In lexicon, this strategy is evident in students' attempts to learn the most frequent words due to their usefulness and practicality" (p. 128). The following incorrect collocations list, taken from the corpus, can be traceable to the strategy of overgeneralization.

Learner-made collocation	Target collocation	Learner-made collocation	Target collocation		
V+P	Play by	Play with	A+N	Alert child	Bright child
V+P	Adapt with	Adapt to	A+N	Awful method	Terrible method
V+P	Accustomed with	Accustomed to	A+P	Good in	Good at
V+P	Quarrel by	Quarrel with	V+N	Make balance	Create balance
P+N	In TV	On TV	V+N	Make a company	Establish a company
P+N	In university	At university	V+N	Technology go up	Technology develop

Table 11 displays the frequency and percentage of collocational errors attributable to intralingual transfer. As the Table shows, the percentage of incorrect collocations affected by intralingual transfer accounts for 30 percent of collocational errors, which is 26.7 percent lower than that of interlingual transfer. As far as the distribution of intralingual collocational errors among two groups is concerned, Azeri learners made 33.8% of intralingual errors, while the percentage for the Persian learners was 26.5%. On the whole, the percentage of overgeneralization in production of English collocations was highest in adverb-adjective category (100%) and lowest (0.0%) in verb-adverb category.

Table 11. Intralingual collocational errors across learners with different L1

Type of Collocation	Persian speakers			Azeri speakers			ΣT	ΣL2	Σ%
	T	L2	%	T	L2	%			
1 Verb+ Preposition	22	2	9.0	19	4	21.0	41	6	14.6
2 Verb+ Noun	24	8	33.3	17	5	29.4	41	13	31.7
3 Noun+ Preposition	4	2	50	11	10	90.9	15	12	80.0
4 Adjective+ Noun	15	4	26.6	13	1	7.7	28	5	17.9
5 Adjective+ Preposition	5	3	60	5	1	20	10	4	40.0
6 Adverb+ Adjective	2	2	100	3	3	100	5	5	100
7 Verb+ Adverb	7	0	0.0	3	0	0.0	10	0	0.0
Total	79	21	26.5	71	24	33.8	150	45	30.0

T= Total number of errors per category

L2= Number of errors due to intralingual transfer

4.3.3 Incorrect collocations due to unfamiliarity with collocation structure

A number of collocational errors identified in the corpus cannot be traced back to interference from L1 or overgeneralization, but rather to a lack of knowledge of the structure of the whole expression. Collocational errors due to unfamiliarity with collocation structure accounted for 13.3% of errors. Some examples of such errors appear in Table 12:

Table 12. Collocational errors due to unfamiliarity with collocation structure

Type	Learner-made collocation	Target collocation
V+P	Pick up the fruit	Pick the fruit
V+P	Come across by	Come across #
V+P	Enjoy about	Enjoy#
V+P	Nag about me	Nag at me
A+P	Safe in (job)	Job security
A+P	Smart in (lessons)	Good at lessons

As can be seen from the examples, these collocational errors cannot be attributed to either negative transfer from the first language or overgeneralization but rather to a lack of familiarity with the structure of the whole expression. Although more remote sources such as being linked to L1 or intralingual transfer may also account for such errors, what seems to be the most plausible reason for most of the items listed above seems to be a lack of familiarity with the target collocation. For example, "safe in" though an acceptable combination in English is not acceptable in the sentence: *He is not safe in his job*. This is the case for "smart in" too which, despite being recorded in the BNC, is unacceptable in the sentence: *I was smart in my*

lessons. Likewise, “enjoy about”, though may be considered as an interlingual error in the sense that it is followed by a preposition, it is not the case, since in Persian the preposition which is used after *enjoy* is “from” not “about.” So such errors can be attributed to the learner’s lack of familiarity with the structure of the target language collocations. As Hussein (1990) rightly points out, the correct production of such collocations in the target language requires prior familiarity with them either through reading or listening.

5. Discussion

5.1 The most challenging types of collocations

Content analysis was carried out in order to provide answers to the research questions. The first question dealt with identifying the most problematic type of collocations for Iranian EFL learners. The corpus analysis revealed that different types of collocations produced different degrees of challenge to EFL learners. Based on the results, prepositional collocations in general and verb-preposition collocations in particular made the most problematic types of collocation for Iranian learners of English. As Table 3 showed, after verb-preposition collocations, the major collocational error categories were related to verb-adverb (25%), verb-noun (23.8%), adjective-preposition (19.6%), adjective-noun (14.7%), noun-preposition (13.8%) and adverb-adjective collocations (5.4%), respectively. Hence, our findings correspond to the previous findings by Delshad (1980), Zarei (2002) and Koosha and Jafarpour (2006) all of whom found that most Iranian EFL learners have difficulty in the use of English preposition-based collocations.

Moreover, as Table 3 displays, the percentage of errors for “adverb-adjective” combinations was lowest among other types of collocations in the corpus. This is not surprising since out of the 92 adverb-adjective combinations produced by the participants, 59 (64%) of the adjectives were modified using the adverb “very” (for example, very naughty, very simple, very clever, very busy, etc.) which is the most common intensifier in English. Such an overuse of “very” by Iranian EFL learners can be traced back to their L1 where they modify most of the adjectives using Persian equivalent of this adverb. On the other hand, the participants’ tendency to overuse “very” results from their lack of lexical competence in English, as they have no other proper intensifier at their disposal such as *terribly*, *extremely*, *highly* and *awfully* to modify the adjectives such as *noisy*, *beautiful*, *enjoyable*, and *boring*, respectively. This finding is in line with that of Fan (2009) who found that Hong Kong learners overused the amplifier “very” compared to British learners. In her study, the intensifying

adverb “very” was used 15 times in the HK corpus but only 7 times in the British corpus.

Furthermore, some instances of the “adjective-noun” combinations as *good grade*, *good technology* and *good accent* used by the participants, though acceptable in English (according to *Oxford Collocations Dictionary for Students of English* and *BNC*), were less common in English than their alternatives. They were found to have only 5, 4 and 1 occurrences in *British National Corpus*, while their alternatives (i.e. *high grade*, *high technology* and *no accent*) yielded 50, 50 and 11 records of frequency, respectively. A total of 23 nouns were modified using the adjective “good”. By the same token, *noisy child*, *big wish*, *men workers* and *messy person* with 2, 1, 1, and 1 occurrence in BNC were among the adjective-noun collocations whose alternatives were found to be *naughty child*, *greatest wish*, *male workers* and *clumsy person* with 11, 10, 33 and 12 records of use respectively. Such instances resulted in the adjective-noun collocations occupying the second rank of the most frequently used collocations and the third rank of the least frequently occurring collocational errors with 14.7% percent.

5.2 The role of L1 in the production of collocations

The second question concerned the extent to which the use of collocations was affected by EFL learners’ first language. As the corpus analysis showed, 56.7 percent of collocational errors made by Iranian EFL learners were attributable to negative transfer from the first language. According to Brown (2000), the beginning stages of second language learning are more susceptible to interlingual transfer. Hence, based on Brown’s claim and high percentage of collocational errors made due to interference from the first language in the current study, it can be concluded that interlingual transfer is a major learning strategy employed by EFL learners, even at intermediate level. The results obtained are in line with those of Hussein (1990), Nesselhauf (2003), Mahmoud (2005), Koosha and Jafarpour (2006), Fan (2009) and Sadeghi (2009) all of whom found that interlingual transfer is a major factor of collocational errors. However, this finding runs counter to the claims made by Dechert and Lennon (1989; as cited in Nesselhauf, 2003) who hypothesized that L1 influence is not very important in making collocational errors.

6. Conclusion

Our findings verified the belief that collocations pose an area of difficulty for the Iranian EFL learners studied in this research project. As the empirical data in the study demonstrated, the participants’ collocational knowledge

and use were less than being satisfactory, hence confirming the claim made by Farghal and Obiedat (1995) that most second language learners' collocational knowledge lagged far behind their vocabulary knowledge. This implies that collocations should be given due attention by EFL learners, teachers and materials developers.

First of all, learners should be convinced of the usefulness of including collocations in their communication activities since it is natural for learners of a foreign language to want to be like a native speaker. Moreover, EFL learners' failure in the correct production of collocations, to some extent, stems from the language teachers' tendency to teach words through definition rather than as parts of their collocational patterns. Actually, abundant collocational errors observed in this study should prompt teachers to reconsider their approach towards vocabulary teaching. It is, therefore, important for language teachers to raise the awareness of L2 learners about this challenging aspect of language and to foster their learning of collocations. According to Woolard (2000), an effective way for improving learners' collocational knowledge is to focus on a selection of students' mis-collocations. By identifying learners' mis-collocations, teachers not only understand the nature of learners' mis-collocations but integrate them into classroom activities at appropriate times in an attempt to improve vocabulary teaching.

Apart from language teachers, materials developers play a significant part in most language programs. They need to revise existing materials so as to take the multi-word units into consideration. As Richards and Renandya (2002) rightly point out, the role and uses of materials in a language program are an important aspect of language curriculum development since they provide a major source of contact with the language for learners.

As to the role of learners' L1 in the production of English collocations, the results revealed that the participants were adversely affected by their mother language in the oral use of collocations, so that around 57% of the collocational errors were due to the negative transfer from the first language. Hence, an immediate implication is a need for the selection and teaching of collocations with reference to learners' L1.

This study, like many other research studies, was not without its limitations. The fact that the study was small scale in nature constitutes the first limitation of the present study. Thus, further research should be conducted on a larger scale by recruiting more homogeneous participants from universities and institutions in different areas in Iran and other EFL contexts to further explore the issues addressed here. Second, in this study the analysis of learners' collocational use and errors was only based on their

oral production, hence there is a possibility of the learners' collocational use and errors caused by factors other than their lack of actual linguistic competence (e.g. distraction or spontaneous as opposed to monitored use of language). Therefore, the investigation of the learners' collocational use and errors in their written production would yield more reliable results and reveal the true picture of their collocational repertoire. Additionally, a comparison can be done across male and female participants in using collocations.

References

- Akbari, R. (1995). *The role of language-specific collocations on the written production of intermediate Iranian students learning English as a foreign language*. Unpublished MA thesis, Tarbiat Modarres University, Tehran, Iran.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formula sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10, 245-261.
- Brown, H. D. (2000). *Principles of language learning and teaching* (4th Ed.). London: Longman.
- Cruse, D. A. (1986). *Lexical semantics*. Cambridge, UK: Cambridge University Press.
- Erdogan, V. (2005). Contribution of error analysis to foreign language teaching. *Mersin University Journal of the Faculty of Education*, 1, 261-270.
- Fan, M. (2009). An exploratory study of collocational use by ESL students: A task based approach. *System*, 37, 110-123.
- Firth, J. (1957). Modes of meaning. In J. Firth (Ed.). *Papers in linguistics* (pp. 130- 158). London: Oxford University Press.
- Harmer, J. (2001). *The practice of English language teaching* (3rd Ed.). London: Longman.
- Hill, J. (2000). Revising priorities: from grammatical failure to collocational success. In M. Lewis (Ed.). *Teaching collocation: further developments in the lexical approach* (pp. 49-50). Hove, UK: Language Teaching Publications.
- Hussein, R. F. (1990). Collocations: The missing link in vocabulary acquisition amongst EFL learners. *Papers and Studies in Contrastive Linguistics*, 26, 123- 136.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27, 4-21.
- James, C. (1998). *Errors in language learning and use: Exploring Error*

- Analysis*. London: Longman. Koosha, M., & Jafarpour, A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal*, 8, 192-209.
- Lewis, M. (2000). *Teaching collocation: Further developments in the lexical approach*. Hove, UK: Language Teaching Publications.
- Li, C. C. (2005). *A study of collocational error types in ESL/EFL college learners' writing*. Unpublished MA thesis, University of Ming Chuan, Taiwan.
- Mahmoud, A. (2005). Collocation errors made by Arab learners of English [Teachers Article]. *Asian EFL Journal*. Retrieved November, 14, 2009 from: http://www.asian-efl-journal.com/pta-August_05_ma.php.
- McCarthy, M. 1990. *Vocabulary*. Oxford: Oxford University Press.
- Nattinger, J. & Decarrico, J. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics Journal*, 24, 223-242.
- Nowruzi Khiabani, M. (2000). Sources of collocational clashes. *Zaban & Adab*, 8, 1- 11.
- Oldin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- Richards, J. C. (1974). *Error analysis: Perspectives on second language acquisition*. London: Longman.
- Richards, J. C., & Renandya, W. A. (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge: Cambridge University Press.
- Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd Ed.). London: Longman.
- Sadeghi, K. (2009). Collocational differences between L1 and L2: Implications for EFL learners and teachers. *TESL Canada Journal*, 26, 100-125.
- Schmitt, N. (2002). *An introduction to applied linguistics*. New York: Oxford University Press.
- Shahheidaripour, G. (2000), *Interlanguage development and collocational clash*. An excerpt from the Author's MA thesis, Azad University of Kerman, Iran.
- Sharpe, P. J. (2009). *Barron's TOEFL* (11th Ed.). New York: Barron's Educational Series.
- Shin, D., & Nation, P. (2008). Beyond single words: the most frequent collocation spoken English. *ELT Journal*, 62, 339-347.

- Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. New York: Palgrave Macmillan.
- William, J. (2000). Testing ESL learners' knowledge of collocations. *ELT Journal*, 35, 115-122.
- Woolard, G. (2000). Collocation- encouraging learner independence. In M. Lewis (Ed.). *Teaching collocation: Further developments in the lexical approach* (pp. 28-46). Hove: Language Teaching Publications.
- Ying, S. (2009). Study on collocations in English writing by Chinese students. *Sino- Us English Teaching*, 6, 25-30.
- Zarei, A. (2002). *Patterns of Iranian advanced learners problems with English collocations*. Unpublished doctoral dissertation, University of Isfahan, Iran.