The Study of Formal and Semantic Errors of Lexis by Persian EFL Learners

Mohammad J. Rezai, Fereshteh Davarpanah

Abstract—Producing a text in a language which is not one's mother tongue can be a demanding task for language learners. Examining lexical errors committed by EFL learners is a challenging area of investigation which can shed light on the process of second language acquisition. Despite the considerable number of investigations into grammatical errors, few studies have tackled formal and semantic errors of lexis committed by EFL learners. The current study aimed at examining Persian learners' formal and semantic errors of lexis in English. To this end, 60 students at three different proficiency levels were asked to write on 10 different topics in 10 separate sessions. Finally, 600 essays written by Persian EFL learners were collected, acting as the corpus of the study. An error taxonomy comprising formal and semantic errors was selected to analyze the corpus. The formal category covered misselection and misformation errors, while the semantic errors were classified into lexical, collocational and lexicogrammatical categories. Each category was further classified into subcategories depending on the identified errors. The results showed that there were 2583 errors in the corpus of 9600 words, among which, 2030 formal errors and 553 semantic errors were identified. The most frequent errors in the corpus included formal error commitment (78.6%), which were more prevalent at the advanced level (42.4%). The semantic errors (21.4%) were more frequent at the low intermediate level (40.5%). Among formal errors of lexis, the highest number of errors was devoted to misformation errors (98%), while misselection errors constituted 2% of the errors. Additionally, no significant differences were observed among the three semantic error subcategories, namely collocational, lexical choice and lexicogrammatical. The results of the study can shed light on the challenges faced by EFL learners in the second language acquisition process.

Keywords—Collocational errors, lexical errors, Persian EFL learners, semantic errors.

I. INTRODUCTION

ERROR correction, whether oral or written, is the process of providing clear and comprehensive feedback on students' grammatical deviances for the purpose of improving the students' ability to write accurately. Providing written error correction play an important role in guiding, motivating, and encouraging students to improve their accuracy in L2 writing.

Reference [4] suggests that the second language learning process is not very different from that of the first language, and the feedback a L2 learner gets upon making errors benefits him in developing the L2 knowledge. Reference [12]

M. J. Rezai is a faculty member of English department, Yazd University, Safaiyeh, Yazd, Iran (corresponding author, phone: +98-9132578649; e-mail: mrezai@ yazd.ac.ir).

F. Davarpanah was with the English Department, Yazd University, Iran.. She is now with Safir language Institute, Yazd, Iran (e-mail: fereshtedavarpana697138@gmail.com).

emphasized that accounting for why an error was made is the most important step in any attempt to understand the process of second language acquisition and to improve the way a second language is learned. The results of the many researches highlight the importance of analyzing the learners' performance. Error analysis may provide a valid guide for language teaching materials preparation and sequencing, valuable diagnostic information concerning learners' progress along the development trajectory leading to target language competence, and help the teachers to understand the process of SLA better. Language teachers can develop a class atmosphere in which the learners expose the language they know to the teacher, and classmates in a way that tends to lower the learner's affective filter. Since 'error' is a necessary factor in the development of the learners towards the target language norm, a class where learners are encouraged to expose their TL knowledge will certainly lead to a better and faster TL acquisition.

The sources of error commitment can be either interlingual or intralingual [4]. While interlingual errors are caused mainly by mother tongue interference, intralingual or developmental errors originate in factors including simplification, overgeneralization, hypercorrection, faulty teaching, fossilization, avoidance, inadequate learning, and false concepts hypothesized. Language learning errors involve all language components: the phonological, the morphological, the lexical, and the syntactic. A lexical error involves inappropriate direct translation from the learner's native language or the use of wrong lexical items in the second language while syntactic errors embrace errors such as word order, subject-verb agreement, and the use of the presumptive pronoun in English relative clauses.

Written errors can benefit language teachers and students. They provide relevance to language teaching, particularly in correcting errors in the classroom, providing grammatical explanations, improving curricula, and developing programs and instructional materials. Language teachers can find a more realistic attitude towards errors. Language learners will realize what they really need to learn and, therefore, can rectify any linguistic deviances they are facing with.

The study of learners' errors can help language teachers to evaluate their teaching practices and their learners' competence in the process of writing. Teachers can also select the writing strategies and topics that are best suited to their students. Competency and proficiency in the language seem indispensable nowadays as the world needs globally competitive individuals, who can fluently and competently use the international language – the English language.

In Iran, there are many language institutes in which English is being taught as a foreign language. In these institutes, students face a lot of problems in vocabulary acquisition i.e. finding the exact equivalent for words and phrases as they are used in first language. The product of their work is not comprehensible enough and there are a lot of errors in their compositions. These errors can be due to transfer from L1, interlingual, or over-generalization of L2 system or intralingual error.

The current study attempted to investigate the kinds and frequencies of formal and lexical errors in the English writings of Persian EFL learners. Furthermore, it examined the errors on three levels of English proficiency (lower intermediate, upper intermediate, and advanced) to follow the students' interlanguage development.

II. LITERATURE REVIEW

Early works in error analysis dealing with L2 data, both spoken and written, focused on error classification and causes of error commitment. In what follows, a number of studies are reviewed highlighting the error commitment causes and frequency of errors.

Reference [6] addressed the lexical errors made by 20 EFL learners of Business English. The study participants were junior learners who had been studying English in Tunisia for eight years. They used Arabic and French at school, but the predominant language among the family members and friends was Arabic. The participants had an age range of 19 to 21 years and did not have much exposure to writing as it was not prioritized in their syllabus. The analysis of data, within the framework of [8], revealed that the rate of formal errors (98.44%) far outweighed that of semantic errors (0.05% of the total lexical errors).

Reference [13] dealt with lexical error analysis. The objectives of their research were to know the lexical error types, the percentage of each lexical error type and the causes of lexical error type found in the descriptive text writing made by the eighth grade students. The result revealed that the students made 86 lexical errors which were classified into 26 errors of misspelling, 22 errors of semantic confusion, 20 errors of calque, 13 errors of misselection and five errors of borrowing. The results on the analysis showed the highest number of lexical errors in the descriptive text writing was the misspelling having 30.23%. The other types were semantic confusion (25.59%), calque (23.26%), misselection (15.11%), and borrowing (5.81%). The causes of errors varied for each type of lexical error based on their sources and major types of error.

Reference [14] specifically addressed the collocation types in the spoken outputs of 30 intermediate learners of English. Their objective was to identify, classify and account for the collocational errors committed by the EFL learners. The results obtained from the analysis of the corpus indicated that the collocations related to prepositions, particularly verb-preposition collocations were the most problematic types of collocation for the EFL learners. It was further revealed that a large proportion of collocational errors (56.7%) can be attributed to the negative L1 transfer, whereas intralingual transfer accounted for only 30% of the incorrect collocations. Furthermore, a content-analysis of the spoken corpus was undertaken in order to describe the possible collocational errors in the oral productions of the EFL learners. To do so, the lexical combinations were identified and extracted from the corpus, totaling 790 collocations. The analysis revealed seven collocational patterns including (a) verb-preposition, (b) noun-preposition, (c) adjective-preposition, (d) adjective-noun, (e) verb-noun, (f) verb-adverb, and (g) adverb-adjective. The first three patterns were classified as "grammatical collocations" whereas the remaining types constituted "lexical collocations".

Reference [2] conducted a study to explore Arab learners' semantic errors in English and to investigate their possible L1 and L2 sources. Thirty essays written by Yemeni University third-level students majoring in English were analyzed. The classification of the identified errors was made on the basis of the three categories of lexical, collocational and lexicogrammatical errors each of which was classified into further categories and subcategories depending on the errors identified.

There were 1388 semantic errors identified in the study. The results showed that omission of letters category scored the highest number of errors, viz. 251, i.e. 18.08% while misselection of a prefix category was the lowest where only 12, i.e. (0.68%) errors were committed. A hierarchy of difficulty was established where formal misformation category represented the highest extreme and formal misselection the lowest one.

As for the sources of the semantic errors committed by Yemeni Arabic-speaking learners, there were two different sources of such errors, namely, L1, i.e. Arabic, and L2, i.e. English. As for L1 based sources, it was found that the sources of these errors varied between the following different strategies such as translating from Arabic, as in the case of some categories in lexical choice and collocation errors, applying Arabic rules to English as in derivativeness, the Arabic sound system as in the case of the absence of /p/ and /v/ in distortion due to spelling errors among others. As for the L2, i.e. English-based sources, the main source was having false conceptions of the way meaning networks work in English as in the case of paraphrasing, collocation and lexicogrammatical choice errors. In addition, L2-based errors were obvious as in the case of misordering, overinclusion in distortion due to spelling, formal misselection and lexicogrammatical errors.

A study conducted by [17] aimed at presenting a comprehensive taxonomy capable of accounting for Saudi EFL students' lexical errors. It examined the types of lexical errors produced by female Saudi students studying English. The analysis of 96 essays produced by the learners led to a total of 718 lexical errors, with an average of 7.48 errors per essay. The highest proportions of error were related to (a) misselection of suffixes, having a frequency of 128 (17.83%) and (b) direct translation from L1 with a raw frequency of 113 (15.74%). Overall, semantic lexical errors (60.45%) exceeded

formal lexical errors (39.55%).

As to the semantic errors, the confusion of sense relations constituted the most frequent category, occurring 285 times (36.69% of total semantic errors). The error types in the category in terms of the order of frequency included (a) direct translation from mother tongue, (b) using words containing inappropriate meaning, (c) near synonyms, (d) using general terms instead of more specific ones, (e) words which deviate meaning, (f) improper co-hyponyms, (g) wrong binary terms, and (h) overly specific terms. The next most frequent category was "collocation errors", which had a similar frequency of occurrence to the category of the "use of inappropriate words".

Reference [9] tried to determine the extent to which Iranian EFL learners' knowledge of collocation of prepositions is affected by their L1. To this end, 200 senior English majors studying at three universities in Shahrekord served as the participants of this study. Comparing collocational errors in the impromptu and public speech tests illustrated interlingual errors to be much higher than intralingual errors. In both tests, interlingual collocational errors turned out to be above 80% unveiling the dominant effect of L1 (Persian) on collocational errors of Iranian EFL learners.

Reference [7] used a more comprehensive error taxonomy based on James (1998), with some additions from Leech's semantics (1981), to analyze Thai third-year university students' English compositions for lexical errors. The analysis revealed that (a) "near synonyms" were the most numerous errors, followed by "preposition partners" and "suffixes", (b) the students had more difficulty with semantics than the forms of words, and (c) the identified sources of errors were mainly from L2 intrinsic difficulty rather than the first language (L1) transfer. Some error types were common (for example, near synonyms, preposition partners and incorrect suffixation), others were relatively infrequent (for example, prefix type, false friends, blending and inappropriate co-hyponyms). This indicates that errors are not evenly distributed across the error-type spectrum; rather, certain error types appear to be particularly problematic.

As far as formal errors are concerned, the formal misselection of words was the most problematic error category in the data (15.33% of all errors), followed closely by intralingual "distortions" (14.56%). L1-influenced errors ("misformations") were less of a problem (6.90%). This finding suggests that the similarity of form and parts of speech (for example, verb, and noun, adjective and so on) remained a serious problem in the Thai students' writings. The possible reason for the formal misselection is lack of knowledge of words in a word family. When the meaning and the spelling of an intended word were acquired, knowledge of the other members of a word family or grammatical patterns may have been incompletely acquired, which might result in a wrong derivative or grammatical pattern being used in a particular context (for example, *the people who live in the country are *honesty* (honest)). Although "distortions" or misspellings were the second most frequent formal errors, they occurred only slightly less frequently (as 14.56% of the total errors)

than "misselections". Thus misspellings were a problem for the Thai students.

Of the four main semantic types of error, "collocation errors" were the most frequent (26.05% of the total), followed by "confusion of sense relations" (24.9%), "stylistic errors" (8.04%) and "connotative meaning" (4.21%). Together, errors in the categories "collocation errors" and "sense relations" accounted for about half of all errors in the study. Overall, semantic errors were roughly twice as frequent as formal errors (63.22% vs. 36.78%).

These results suggest that formal errors were less problematic for the students than semantic errors. However, they made up a considerable percentage of the total errors, and this implies that the students would benefit from developing their morphological and formal knowledge, as [15] suggests.

Reference [11] offered an analysis of lexical errors committed by Arab EFL learners. He paid particular attention to the study of collocation errors. The researcher presented empirical data verifying the informal observations and theoretical assertions that EFL learners produced inappropriate lexical combinations. The Arabic-speaking participants, who were undergraduate students of English, produced 42 essays. The data analysis revealed a total of 420 collocations, out of which, 64% were erroneous. The majority of collocational errors were lexical rather than grammatical (80% vs. 20%). The researcher attributed 61% of the deviant combinations to negative transfer from L1, i.e. Arabic.

Reference [20] examined the lexico-semantic relation errors in ESL writing. The data were taken from students' letter texts in secondary schools randomly selected from two states. The lexico-semantic errors were traced to four linguistic sources: collocation, generalization, similarity, and duplication. From the data, seven sub-categories of errors were identified. The findings revealed that collocation errors were predominant, accounting for 56.5% of the total lexico-semantic errors.

Reference [1] analyzed the lexical errors in the written English of 110 Saudi female college freshman students. Like Sheshsha (1993), she classified her errors into two main types, intralingual and interlingual. The intralingual errors comprised eight categories along with a further category for uninterpretable errors. The interlingual errors, on the other hand, included literal translation and divergence, similar to what was observed by [18]. She agreed with [18] that intralingual errors were more frequent than the interlingual ones. However, she stated that lexical interference in one particular language or between the two languages can be regarded as the major source of lexical errors. Al-Jabri also acknowledged that the categories of lexical errors were neither exclusive nor discrete.

Reference [18] conducted a study on lexical errors committed by 48 Saudi undergraduate students majoring in English. The results of the study yielded five categories of errors including (a) confusion of words having formal similarities, (b) confusion of words having similar meaning, (c) unnatural collocation, (d) literal translation, and (e) divergence. All the error categories were then divided into either interlingual or intralingual error types. The results of the

study showed that unnatural collocation errors were the most frequent (38.71%), while literal translation errors were ranked as the second (23.65%). The least frequent error type in the analysis was related to the confusion of words having similar meaning, which covered 11.29% of the total errors. Finally, [18] concluded that intralingual errors, which have their source in the target language, were more frequent than interlingual errors.

Reference [21], analyzing the writings of Arab students and working with quite a comprehensive lexical error taxonomy, argued that determining a taxonomy of error types committed by EFL writers can shed light on the nature of interlanguage and can illuminate lexical choice strategies adopted by EFL learners. He offered a typology of 13 lexical errors. According to his findings, the most frequent lexical error was assumed synonymity, which covered 23.5% of the total errors. He pointed out that a majority of lexical choice errors can be caused by L1 interference or negative transfer which can be manifested through assumed synonymity, literal translation, derivativeness and idiomaticity.

III. METHODOLOGY

A. Research Questions

In this study, three types of categorical variables were studied: (a) two levels of formal errors (misselection and misformation); (b) semantic errors within three levels (lexical choice, collocations, lexicogrammatical choice); and (c) three levels of proficiency (low-intermediate, upper-intermediate and advanced).

Accordingly, the research questions under investigation in this study included the following:
1- What types of semantic errors are common in the writings of EFL learners?
2- What types of formal errors are common in the writings of EFL learners?
3- What role does the proficiency level play in error commitment?

B. Research Design

The method used in this study was quantitative. The data in quantitative research are gathered in a numerical fashion. The quantitative investigation method was chosen since the purpose was to look for the association between proficiency levels and error commitment through analyzing numerical data with chart and figure representation. This approach allows the researcher to examine the relationship between two variables within population, focuses on gathering numerical data, often arranged in tables, charts and figures, and generalizes the results across the groups of people.

The research design of this study was further categorized under analytic and deductive approaches. Reference [16] defines analytic approach as one which identifies a single factor or a cluster of factors. It focuses on a more specific aspect of language proficiency. An analytic approach means that the second language phenomenon is analyzed in terms of its basic ingredients and a single or a cluster of constituent parts is scrutinized, leaving the other constituent factor aside.

Given the fact that it would take so long to collect data from the entire population, stratified random sampling was utilized in the current study in order to have control over the participants, to ensure that the result of the study can be reasonably extended to the whole population, to be protected from obtaining extreme result of large population, and to increase the efficiency of the result. To do so, the whole population of adult learners, 12 classes, were grouped into three proficiency levels of lower intermediate, upper intermediate and advanced, out of which, half of the learners of the three proficiency levels, as the representative of the whole population, were selected.

The data of this work were gathered, organized and drafted in Safir language academy, Yazd branch. Data collection, data analysis, and deduction of final results began in April 2018 and ended in September 2018. An oral interview was used to place the learners into their proficiency level in line with the norms of the language institute.

In this study, 60 adult Iranian English learners across three proficiency levels with various age groups, from 20 to 35, were chosen as samples of the research. The selected sample consisted of 15 intermediate learners, 30 advanced learners, and 15 low intermediate female learners all of whom voluntarily participated in the study.

C. Analysis of Collected Data

Statistically, there were a number of categorical variables including: (a) proficiency with three levels: low intermediate, upper intermediate, and advanced, (b) errors of lexis with two levels: formal and semantic, (c) formal errors with two levels, each of which with related sublevels: formal misformation and formal misselection, and (e) semantic errors with three levels: lexical choice, collocation, and lexicogrammatical.

To analyze the obtained data, a taxonomic framework was adopted combining [8] and [2] error categories. Based on the given framework below, all the learners' deviances were identified and classified.

1. **Formal Errors**
I.   Formal Misselection
(a) Misselection of a Prefix
(b) Misselection of a Suffix
II. Formal Misformations
(a) Direct Translation from L1
(c) Borrowing
(d) Coinage
(e) Distortion due to Spelling
(f) Calque
2. **Semantic Errors**
I.   Lexical Choice
(a) Assumed Synonymy
(b) Derivativeness
(c) Confusion of Binary Terms
(d) Paraphrase
(e) Idiomaticity
(g) Similar Forms
II. Collocations

(a) Collocate Choice
(b) Contextualization
(c) Wrong Forms
III. Lexicogrammatical Choice
(a) Adjectives in place of Nouns
(b) Nouns in place of Adjectives
(c) Adjectives in place Adverbs
(d) Adverbs in place of Adjectives
(e) Verb in place of Noun
(f) Noun in place of Verb
(g) Verb in place of Adjective
(h) Adjective in place of Verb

SPSS 22.0 software was used to analyze the collected data. Since the research dealt with the frequency of different types of errors, two types of chi-square tests were used. For variables with one category, a chi-square test for goodness of fit was used, while a chi-square test for independence was used for analyses with two or more categories.

IV. RESULTS AND DISCUSSION

The purpose of the study was to investigate the written formal and semantic errors of lexis. A corpus of 9600 words, consisting of 2583 errors, was analyzed in order to determine the precise frequency of formal and semantic errors of lexis within three different proficiency levels. To this end, a combination of [8] and [12] taxonomy was followed to classify and quantify the number of lexical errors in the students' writings.

A. Analysis of Formal and Semantic Errors of Lexis

Table I presents the total frequency of lexical errors observed in the corpus. Among the distribution of 2583 errors, 2030 formal and 553 semantic errors were obtained. In other words, 78.6% of errors were formal lexical, while 21.4% belonged to the semantic category. As it is shown, formal errors of lexis (78.6%) are more frequent than semantic errors (21.4%) in the corpus.

TABLE I
FREQUENCY OF FORMAL AND SEMANTIC ERRORS

Lexical Errors	Frequency	Percent
Formal	2030	78.6
Semantic	553	21.4
Total	2583	100.0

B. Analysis of Formal Errors

The result of the analysis revealed that among formal error categories, misformation errors (98%) were much more frequent than misselection errors (2%). Among misformation subcategories, spelling distortion errors (84.9%) and calque (13.4%) were more frequent than other subcategories.

Considering the related subcategories, misselection of suffix (53.7%) was more frequent than misselection of prefix (43.9%). The errors related to the misselection of suffix were more frequent at low intermediate level (24.4%) compared to the advanced and the upper intermediate level (14.6%). This suggests that the students had more problems in using words with the right word classes or derivative forms (for example,

noun, verb, adjective and adverb) than the similarity of forms within the same class. One source of suffix errors can potentially be attributed to the wide use of bilingual dictionaries, particularly pocket dictionaries. Though the particular word class is given, there is no explanation or examples of how to use these word classes in context. The students, therefore, tend to memorize a core word or a frequently used one and apply it in every context, even where it is not the one required. Potentially, the wrong use of word class is grammatically driven. However, nouns, verbs, adjectives and adverbs are also lexical words, which deal with word structure and the constraints of word structure can result in suffix errors.

The results showed that the misformation errors were also more frequent at advanced (41.5%) and low intermediate levels (37.4%) compared to the upper intermediate learners (19.01%). It seems that the upper intermediate learners act more conservatively when faced with the selection of appropriate affixation. The findings of the present study reveal that while calque is more frequent at the low intermediate level (7.1%), spelling distortion errors are more frequent at the advanced level (37.5%).

Among spelling distortion subcategories, the most frequent errors included omission followed by misselection, overinclusion and misordering across all the three proficiency levels.

All the spelling distortion subcategories in the current study were more frequent at the advanced level in comparison with low and upper intermediate levels, among which, omission errors were the most frequent ones followed by misselection, overinclusion and misordering. One reason for the higher incidence of distortion errors is that the advanced learners have a larger number of words in their lexical repertoire and use them more frequently than the lower proficiency learners.

C. Analysis of Semantic Errors

The obtained results in the current study showed that across the semantic error categories, the L2 learners committed similar frequencies of errors across the relevant subcategories: collocation (33.6%), lexical choice (33.1%), and lexicogrammatical (32.7%). The participants' data revealed that the lexical collocational errors slightly outnumbered the other semantic subcategories.

Wrong collocation is one of the main sources of lexical errors made by Persian learners of English. Collocation seems to be a language-specific phenomenon, i.e. each language appears to have its own collocation patterns although some of these might be similar in two or more languages. The wrong choice of collocation (see points 1 and 2 below) can be attributed to translation from Persian into English and to the dependence on monolingual dictionaries that offer one word synonym without explanation or examples.
1. *I would like to *take decision by myself (make).*
2. *Society is going to deprive you *from freedom (of).*
The results of the present study indicated that errors of lexical choice were more frequent at the low intermediate level (82.1%), while the collocation errors (see the examples)

were more frequent at the upper intermediate level (26.2%).

Nonetheless, lexicogrammatical errors of lexis were the most prevalent category across the advanced level (39.2%). Considering lexical choice subcategories, errors of *similar forms* (59.7%) and *assumed synonymy* (22.4%) were more frequent than other subcategories. Examples of similar forms errors are illustrated in 3-5, while examples of assumed synonymy errors are given in 6-7 below:

3. *I like to buy technology *staff like headphones (stuff).*
4. *She always *advice us (advise).*
5. *It badly *effects us (affect).*
6. *He likes his *work (job).*
7. *It is not *related to this topic (relevant).*

According to the results of the present study, *assumed synonymy* errors were more frequent at upper intermediate level (13.9%) than advanced (5.2%) and low intermediate level (3.6%). The errors related to the confusion of binary terms were more frequent at low intermediate level (5.7%). Similar forms were more frequent at low intermediate level (26.3%) than advanced (20.6%) and upper intermediate (13.4%) levels. It seems the students who make *similar form* lexical error type already know enough English vocabulary, but they do not know the usage of the words in the sentences. In other words, they possess an acceptable breadth of vocabulary; however, they lack an adequate depth of vocabulary knowledge. Additionally, they know two or more English words which have the same meaning.

D. Analysis of Formal and Semantic Errors across Proficiency

The obtained results in the current study reveal that errors of lexis, across the proficiency levels, are more frequent at the advanced (40.5%) and low intermediate (38.7 %) levels in comparison with the upper intermediate level (20.8%). Different causes such as lexicon organization, lexical repertoire, and compensatory strategy use might account for this varied nature of lexical error production.

Formal errors were more frequent at the advanced (33.3%) and the low intermediate level (30%) in comparison with the upper intermediate level (19.5%). The semantic errors, on the other hand, were more frequent at low intermediate (8.7%) and the advanced level (7.2%) than the upper intermediate level (5.5%).

One major reason for the above observation is that at this developmental interlanguage state proficiency is associated with grammatical accuracy or fluency instead of lexical correctness. As indicated by [10], some factors such as inadequate learning of general writing strategies and transfer of literacy from L1 can act as deterring factors toward the mastery of writing. Learners at the advanced level of language knowledge continue to make lexical errors of all types. The fact that advanced learners of EFL are still of limited linguistic and lexical competence may also account for this error production.

The results suggest that formal errors are less problematic for the students than semantic errors at the low intermediate level. However, they make up a considerable percentage of the total errors, and this implies that the students would benefit from developing their morphological and formal knowledge, as [15] suggest. Semantic knowledge might be more difficult to acquire, since various word knowledge facets are required. It can be argued that these aspects are among the more advanced elements of word knowledge, and that learners are not likely to master them until later in the acquisition process.

The learners across the three proficiency levels committed more formal errors than the semantic ones although the extent was less for the learners at the low intermediate level. Formal errors of lexis which were the most frequent errors in the corpus, were more frequent at the advanced level (33.3%), while the semantic errors were more frequent at the low intermediate level (8.7%).

Considering the fact that the misselection errors were more frequent at the advanced level (0.9%), among subcategories of misselection errors, misselection of suffix was more frequent at the low intermediate level (24.4%), and misselection of prefix was more frequent at the advanced level (29.6%).

Based on the findings of the present study, misformation errors (41.5%) and among misformation error subcategories, spelling distortion (37.5%) were more common at the advanced level. Among spelling distortion subcategories, omission (19.1%), misselection (11.9%), overinclusion (9.1%), and misordering (3.3%) were more frequent at the advanced level.

It seems that at this stage, the students have a variety of related forms floating in their heads as well as several "holes" in their interlanguage system. When a sign is required for which they have no form, students create new words based on patterns found in their own L1 or in the target language itself.

Among formal errors subcategories misformation errors (37.4%) were more common than misselection errors (0.7%). Among semantic error categories, lexical choice errors (33.3%) were more frequent at low intermediate level.

The errors of lexicogrammatical choice, on the other hand, were the most frequent (32.9%) subcategory of semantic errors at the advanced level. Persian EFL learners do not seem to be able to distinguish between base adjectives and derived adjectives on the one hand and adjectives and nouns, on the other.

According to [19], learners overextend the previously learned words leading to semantic confusion. Such a process of overextension is a typical behavior observed by children in their initial stages of language acquisition. In many cases, the semantic confusion does not derive from a failure in choosing the correct word between two, but rather from the use of the only word the learner knows. The lexical item selected by the learners enjoys a similarity to the intended item. As a result, we expect higher levels of lexical competence on the part of L2 learners as they get more and more proficient. Learners gradually gain higher levels of lexical accuracy, leading to a transparency of the association between meaning and concept, and consequently, confusions get reduced.

In case of *similar forms* errors, it seems that participants drag from their passive lexical repertoire a word that shares the same phoneme/grapheme thinking that he has the correct

choice. These errors differ in one or two phonemes/ graphemes.

Sometimes, the second language learners are unable to differentiate various functions of some items that look similar. Some of the words look similar either on account of semantic affinity or due to common source or etymology or other common features. They create confusion and uncertainty to the ESL learners. It is quite possible that the learners are not able to bring out inappropriate word formation, despite their cognitive understanding of the stem form of a lexical item.

V. Concluding Remarks

The current study is contributive to EFL research since it has shed light on the lexical errors made by EFL learners in the Persian context. It has further revealed the difficulties faced by EFL learners in learning English lexis.

Investigation of the related data in the current study reveals that among the distribution of errors (2853), 78.6% of errors were formal lexical, while 21.4% belonged to the semantic category. The results clearly indicate that formal errors of lexis were more frequent than semantic errors in the corpus. It can be seen that different proficiency stages are characterized by different error types [10]. Thus, an increase in the production of certain types of lexical errors is a sign of development and not of lack of progress in acquisition [5]. The results obtained from the study are in line with [6] who suggests that EFL learners face more difficulties with formal than with semantic errors.

The primary purpose of the current study was to explore the most frequent formal errors types. Among formal errors of lexis, misformation errors (98%) were more common than misselection errors (2%). As to addressing the second research question in line with the most frequent semantic errors types, the results revealed that the errors of *similar forms* and *assumed synonymy* were more frequent than the other categories of semantic errors.

To sum up, committing errors in general and lexical errors in particular is an indispensable part of second language acquisition. Therefore, according to [3], we can expect lexical errors by the L2 learners in the process of acquiring the target language. Mastering lexical items is associated with some initial mismatches between form and concept. In this sense, one key distinction among language proficiency levels can be ascribed to the notion of lexical accuracy.

References

[1] Al-Jabri, Samia M. H. (1998). *An Analysis of Lexical Errors in Written English of Saudi College Freshman Female Students.* An unpublished master's thesis, Girls College of Education, Makkah, Saudi Arabia.
[2] Al-Shormani, M. Q. and Al-Sohbani Y. A., (2012). Semantic errors committed by Yemeni university learners: classifications and sources. *International Journal of English Linguistics Analysis.* London: Longman.
[3] Agustin Liach, P.M. (2015). Lexical errors in writing at the end of primary and secondary education: Description and pedagogical implications. *Porta Linguarrum*, 23, 109-124.
[4] Brown, H. D. (2002). *Principles of language learning and teaching.* White Plains, NY: Longman.
[5] Fernández, S. (1997). *Interlengua y análisis de errores en el aprendizaje del español como lengua extranjera.* Madrid: Edelsa.
[6] Hamdi, S. (2016). An analysis of lexical errors in the English compositions of EFL Tunisian learners. *International Journal of Humanities and Cultural Studies, 2*(4), 643-652.
[7] Hemchua, S. and Schmitt, N. (2006). An Analysis of Lexical Errors in the English Compositions of Thai Learners. *Prospect, 21*(3), 3-25.
[8] James, C. (1998). *Errors in Language Learning and Use: Exploring Error Analysis.* London: Longman.
[9] Koosha, M., & Jafarpour, A. (2006). Data-driven learning and teaching collocation of prepositions: The case of Iranian EFL adult learners. *Asian EFL Journal, 8*(8), 64-82.
[10] Lasagabaster, D. and Doiz, A. (2003). *"Maturational Constraints on Foreign-language Written Production",* in M.P. García Mayo and M.L. García Lecumberri (eds.), *Age and the Acquisition of English as a Foreign Language.* Clevedon: Multilingual Matters, 136-160.
[11] Mahmoud, A. (2005). Collocation errors made by Arab learners of English. *Asian EFL Journal,* available at http://www.asian-efl-journal.com
[12] Muriel, Saville-Troike (2006). *Introducing Second Language Acquisition.* Cambridge. Cambridge University Press.
[13] Rochmaniyah, I.A., Suharjito, B., Adi Andayani, M. (2013). An Analysis of Lexical Error in the Descriptive Text Writing Produced by the Eighth Grade Students of MTs. Wahid Hasyim Kunir. *International Seminar and Workshop on ELT.*
[14] Sadeghi, K. & Panahifra, F. (2013). A Corpus-based Analysis of Collocational Errors in the Iranian EFL Learners' Oral Production. *The Journal of Teaching Language Skills, (4)*, 53-78.
[15] Schmitt, N., Zimmerman, C. B. (2002). Derivative word forms: What do learners know? *TESOL Quarterly, 36*(2), 145–171.
[16] Seliger, H. W., & Shohamy, E. (1989). *Second language research methods.* Oxford, UK: Oxford University Press.
[17] Shalaby, N., Yahya, N., & El-Kom, M. (2009). Analysis of Lexical Errors in Saudi College Students' Compositions. *Journal of the Saudi Association of Languages and Translation. 2*(3), 65-93.
[18] Sheshsha, Jamal A. (1993). Lexical Error Analysis in Learning English as a Foreign Language. Social Science Research Series. *Umm Al-Qura Univesity, Makkah, Saudi Arabia, 24,* 5-30.
[19] Singleton, D. (1999). *Exploring the Second Language Mental Lexicon.* Cambridge: CUP.
[20] Taiwo, R. (2001). Lexico-Semantic Relations Errors in Senior Secondary School Students' Writing. *Nordic Journal of African Studies, 10*(3), 366-373.
[21] Zughoul, M. R. (1991). Lexical choice: Towards writing problematic word lists. *International Review of Applied Linguistics, 29,* 45–60.