Measuring the Lexical Richness of Productive Vocabulary in Iranian EFL University Students' Writing Performance

Nazli Azodi

Department of Foreign Languages, Khorasgan (Isfahan) Branch, Islamic Azad University, Isfahan, Iran

Fatemeh Karimi

Department of Foreign Languages, Khorasgan (Isfahan) Branch, Islamic Azad University, Isfahan, Iran

Ramin Vaezi (Corresponding Author)

Department of Foreign Languages, Khorasgan (Isfahan) Branch, Islamic Azad University, Isfahan, Iran

Abstract—This quantitative study aimed at measuring L2 lexical richness of productive vocabulary in the written production of Iranian EFL university students. In fact, this work was conducted to find out how rich the productive vocabulary of Iranian EFL university students is. Productive vocabulary knowledge is to express meaning through speaking or writing, retrieve and produce the appropriate spoken or written word form (Nation, 2001). Considering the importance of size and level of this knowledge among language learners, 210 writing samples such as final examinations and homework assignments of EFL undergraduate students majoring translation at BA level in semesters 2, 4, 6, and 8 at Islamic Azad University, Khorasgan branch, were analyzed by LFP (a tool to measure lexical richness in written texts). The generated profiles matched with available vocabulary frequency lists. The results of this work indicated an increase in the students' ability to use their lexical knowledge throughout years of study at university. However, this knowledge is under the suggested threshold proposed by recent studies. The findings of the research can shed light on the relationship between vocabulary growth and different input conditions. The results of this work can provide insights for language teachers and program planners in developing appropriate materials for enriching vocabulary among learners more effectively.

Index Terms—lexical richness, productive vocabulary knowledge

I. INTRODUCTION

A. Overview

It is believed that vocabulary is the key component of language proficiency and provides much of the basis for how well learners speak, listen, read, and write. It is also claimed that without an extensive vocabulary and strategies for acquiring new vocabulary, learners often achieve less than their potential and maybe discouraged from making use of language learning opportunities around them (Richards &Renandya, 2002).

Second language learners who know less than 1000 word families have difficulty in comprehending and understanding the inferences of a normal text. The same problem exists when they listen to a speech. Even though many words can be recognized but many of them cannot be recalled quickly enough to enable fluent comprehension.

There are two major areas of interest in assessing the vocabulary knowledge of learners. The first one is assessing vocabulary size of learners which is referred to as assessing breadth of knowledge, and the second one is assessing quality of vocabulary knowledge which is referred to as assessing depth of knowledge (Anderson &Freebody, 1981).

Most of the studies have been based on the amount of vocabulary a second language learner needs. It is suggested that although a language makes use of a large number of vocabularies, not all of these vocabularies are equally useful. Some researchers define usefulness in word frequency, that is, how often the word occurs in normal use of the language. The most frequent 10 words account for around 25% of the running words in spoken and written use. The most frequent 1000 words account for around 75% of the running words in formal written texts. Therefore, it is essential to choose carefully what words to pay attention. For these reasons a distinction is made between the high frequency words of a language (Laufer& Nation, 1999).

We all have the experience of being able to understand a word when we see it in a text or hear it in a situation, but not being able to use it in producing language. This shows that there are different degrees of knowing a word. Receptive vocabulary knowledge means being able to recognize one of the aspects of knowledge through reading or listening, and productive knowledge means being able to use it in speaking or writing. Therefore, it is clear that without knowing a basic vocabulary there is little we can do. In this study the ability of Iranian EFL students to utilize their vocabulary

knowledge is on the focus. The attempt is to measure their vocabulary size during their education and examine if their language proficiency is compatible with their productive vocabulary knowledge.

As it was expected EFL students possess a high level of lexical richness at the end of their B.A. courses comparing with early stages of the education. It was intended to study whether it is true or not, and if it is to what extent they can utilize them efficiently. This investigation attempted to find out if lexical richness in its productive point of view changes with the increase of proficiency.

B. Statement of the Problem

The majority of students learning a second or foreign language considers vocabulary as their major priority and agrees that most of their difficulties in the receptive and productive use of language arise from their inadequate vocabulary knowledge (Richards, 1990).

A learner might be able to understand a word when he sees the word but he is not necessarily able to utilize it in speaking or writing. To use the word productively a lot more knowledge of the word is needed.

Finding the vocabulary size of EFL students in different levels of language proficiency, this study attempts to examine the EFL students' lexical richness of their productive vocabulary knowledge and to find out whether this knowledge increases by passing courses at university. As vocabulary sizeitself forms a part of language proficiency, we wanted to see how it changes by reaching the higher levels of education.

As it was expected such learners should be able to utilize the vocabularies they have learned as they reach the end of their education. The attempt was based on such expectations to find out if these learners are able to fulfill it or not.

C. Research Question

This study attempts to find the answer to this question:
How rich is the productive vocabulary profile of Iranian EFL university students?

D. Significance of the Study

Lack of learners' knowledge of highly frequent words is viewed an indication of small vocabulary size. In this case, Nation (2005) proposes that the most frequent words are essential for any real language use.

On the other hand, Laufer and Nation (1995) believed that use of low frequency words is an indicator of richness in a learner's vocabulary which is essential for academic success. This group of vocabulary should be practiced autonomously and indirectly.

The results of this work help teachers to decide where and how learners should be given help with vocabulary learning. And it is also helpful for language planners and administrators to choose appropriate resources with consideration of learners' level of proficiency and their vocabulary needs.

II. REVIEW OF THE RELATED LITERATURE

A. Vocabulary Teaching

For many years vocabulary was the poor relation of language teaching. Its neglect has been in part due to a specialization in linguistic research on syntax and phonology which may have caused vocabulary to be felt a less important element in learning a second language. According to Meara (2005) second language vocabulary acquisition is no longer a neglected area, and vocabulary has been given the pride of place in language teaching methodologies.

In the early stages of learning a language a range of techniques to aid memorization is necessary. In particular, teaching techniques which foster imagistic association should be given to phonological patterns to aid retention in the lexical store. Teaching vocabulary in early language learning requires constant reference to the notion of certain words begin more core than others (Carter, 1998).

According to Nation (2001), vocabulary building strategies should:
- involve choice, that is, there are several strategies to choose from
- be complex, that is, there are several steps to learn
- require knowledge and benefit from training
- increase the efficiency of vocabulary learning and vocabulary use

B. Lexical Frequency Profiles

The lexical frequency profiles (LFP) is a technique to estimate productive vocabulary size explained in Laufer and Nation (1995). LFP was originallyoriginated by Nation as an approach of evaluating whether a particular text is appropriate to be used with learners at a specified level of proficiency. In its easiest form, LFP takes a text as raw input, and outputs a profile that describes the lexical content of the text in terms of frequency bands (Meara, 2005).

LFP analyzes a piece of writing in terms of the proportion of frequent and non-frequent vocabulary in it, which is the percentage of words used from different vocabulary frequency levels (Laufer, 2005).

In Nation's original formulation of LFP (Nation &Heatley, 1996) the bands are explained as follows: the first [band] contains the most frequent 1000 words of English. The second [band] contains the second 1000 most frequent words, and the third [band] contains words not in the first 2000 words of English but which are frequent in upper secondary

school and university texts from a wide range of subjects. All of these base lists contain the base forms of words and derived forms.

Laufer and Nation (1995) claimed that the texts produced by learners at different levels of proficiency have characteristically different profiles, and that the text profiles change in predictable ways as learners' vocabularies grow. This is an attractive idea for vocabulary research, which is seriously short of good tools to assess productive vocabulary abilities, and the claims that Laufer and Nation make for LFP suggest that it needs to be taken very seriously indeed. These claims are that LFP:

- is a reliable and valid measure of lexical use in writing;
- provides similar stable results for two pieces of writing by one person;
- discriminates between learners at different proficiency levels;
- correlates with an independent measure of vocabulary knowledge;
- is a useful diagnostic test;
- is a sensitive research tool (Laufer& Nation,1995; Meara, 2005).

In addition, LFP has been utilized by a number of other authors, includingCobb and Horst (2002), Meara, Lightbown, and Halter (1997), and Muncie (2002). According to Meara (2005), Laufer and Nation are absolutely right to identify the assessment of productive vocabulary as a crucial element in developing theories of vocabulary development in L2 learners.

C. Measuring Lexical Richness

The general term that is used for the characteristics measured by statistics is *lexical richness*. Lexical richness has already been investigated both in *applied linguistics* and in the context of *word frequency distributions*. In *applied linguistics*, a number of tests have been developed for measuring the lexical usage of children or second language learners (Read, 2000). Lexical richness measures are used to evaluate the lexical proficiency level of a child or student, comparing their lexical richness with an external reference point. The most widespread tests are based on the concept of vocabulary diversity, which is assessed using a type-token ratio (TTR) or a TTR-based measure.

In the context of *word frequency distributions*, lexical richness has also been studied. Interestingly, Baayen (2001) has indicated that all mathematical TTR transformations proposed (including the Index of Guiraud, Herdan&Uber) are text length dependent. To be more specific, they are unable to capture the specific structure of the lexicon, which is characterized by a *Large Number of Rare Events* (LNRE): while a small number of words are very frequent, the majority of words occur only a few times, even in large token samples.

More recent work by Herdina and Jessner (2002) has produced a dynamic systems approach to multilingualism, which appears at first sight to have considerable implications for the study of lexicons.

A small number of researches have dealt specifically with developing properties of lexicons. The most important source here is the work of Luc Steels and his colleagues (e.g. Steels, 1995, 1996; Steels & Kaplan 2002) who have reported a series of research studies in which autonomous agents acquire a vocabulary by interacting with each other.

Thus, this study attempts to provide a profile of productive vocabulary in Iranian EFL students. Reid(1993) reports writing demands a rich and appropriate amount of vocabulary, and usually more than 40% of the words in written texts are lexical (content words), whereas in spoken texts the percentage is generally below 40 percent. It reflects the fact that information and ideas are presented in a more concentrated way in written language than is typical of speech. Therefore, here student's writing has chosen to be examined.

III. METHODOLOGY

A. Participants

This study involved a group of Persian learners of English who were in the second, fourth, sixth, and eighth semesters, majoring in translation at Islamic Azad University, Khorasgan Branch. The selection of this university was made since it was convenient and easy to access.

To carry this study, term papers, exam papers, and compositions of the students were used. In total, 210 student papers were manipulated. These students in each semester were from the same university, the same department, and were taught by the same university instructors. Thus, each group of students was as closely matched as possible with regard to input conditions at university. In sum, all the four groups were matched on all variables except the additional year of study. They were all native Persian speakers and their age ranged from 19 to 26.

At this work to get the required result, a cluster design was manipulated that is where research subjects are not sampled independently, but in a group. As this study required samples of student writing, therefore term papers, exam papers, and compositions were used.

As a result in each semester those courses that involve writing were selected and all students' papers and compositions were taken. Therefore, 210 papers were utilized from students in the second, forth, sixth, and eighth semesters.

B. Materials

To run this study, term papers and compositions of the students were used. This research aimed at measuring the richness of productive vocabulary in Iranian EFL university students' writing performance to see whether or not they add to their vocabulary portfolio through years of study at university.

As word choice and proficiency are two skills playing important role in writing, this study analyzed students' writing. For this purpose those courses that involve writing were selected. In semester 2 compositions of students attending Essay Writing classes were analyzed. It should be mentioned that the length of these compositions were not sufficient. Therefore, we asked for their compositions twice. In semesters 4, 6, and 8 courses like Advanced Writing, Contrastive Analysis, and Research Methodology were selected respectively.

C. Instrumentations

In this work first we needed to know how much lexical items learners know in each semester. For this purpose, there is a computer program developed by Laufer and Nation in 1995. They developed a procedure which categorizes the words in a learner's text. It is a tool which attempts to measure free productive vocabulary in the compositions of second language learners.

They called this analysis the lexical frequency profile (LFP) of the text. The LFP analyser program, (now renamed *RANGE*), indicates the numbers and percentages of words and word families in a target English text coming from each of the 3 word lists, plus those which are not recognized. It is basically a measure of lexical use in writing.

In this study, the software *RANGE* was used to analyze the body gathered from students. *RANGE* is used to compare the vocabulary of up to 32 different texts at the same time. For each word in the texts, it provides a range or distribution figure (how many texts the word occurs in), a headword frequency figure (the total number of times the actual headword type appears in all the texts), a family frequency figure (the total number of times the word and its family members occur in all the texts), and a frequency figure for each of the texts the word occurs in. It can be used to find the coverage of a text by certain word lists, create word lists based on frequency and range, and to discover shared and unique vocabulary in several pieces of writing.

RANGE can be used to compare a text against vocabulary lists to see what words in the text are and are not in the lists, and to see what percentage of the items in the text are covered by the lists. It can also be used to compare the vocabulary of two texts to see how much of the same vocabulary they use and where their vocabulary differs.

It is beneficial for example for seeing what low frequency words are in an exam question paper, a technical information note or a text aimed at foreign readers. It may also be used to check the vocabulary of simplified reading texts or language course books to see how many of the words in the texts are among the high frequency words of English. It may also be used to see how much learning the vocabulary of one text helps with dealing with the words in a different text.

RANGE provides a table which shows how much coverage of a text each of the three base lists provides. Table 1 is a sample table of RANGE output:

TABLE 1:
RANGE SAMPLE OUTPUT TABLE

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
one	54/72.0	34/69.4	33
two	2/ 2.7	2/ 4.1	2
three	14/18.7	9/18.4	9
Not in the lists	5/ 6.7	4/ 8.2	????
Total	75	49	44

This depicts that 54 of the running words in the text are in base list one and these 54 words make up 72% of the total running words in the text. In the word list column, one, two, three refer to each of the base lists.

Three ready-made base lists are available. The first (BASEWRD1.txt) includes the most frequent 1000 words of English. The second (BASEWRD2.txt) includes the 2nd 1000 most frequent words, and the third (BASEWRD3.txt) includes words not in the first 2000 words of English but which are frequent in upper secondary school and university texts from a wide range of subjects. All of these base lists include the base forms of words and derived forms. The first 1000 words thus consist of around 4000 forms or types. The sources of these lists are A General Service List of English Words by Michael West (Longman, London 1953) for the first 2000 words, and The Academic Word List by Coxhead (1998, 2000) containing 570 word families. The first thousand words of A General Service List of English Words are usually those in the list with a frequency higher than 332 occurrences per 5 million words, plus months, days of the week, numbers, titles (Mr, Mrs, Miss, Ms, Mister), and frequent greetings (Hello, Hi, etc).

D. Procedure

Participants (N=210) in this study were at the second (N=90), forth (N=40), sixth (N=40) and eighth (N=40) semesters of their English studies at Khorasgan Azad University. In sum, there were four groups of students in this research. This study conducted to measure the productive vocabulary of these students. To do so, we needed samples of their writings.

These samples have taken from those courses that involve writing during years of education at university. To this end courses such as, Advance Writing, Essay Writing, Contrastive Analysis, and Research Methodology selected. For each group, that here means each semester, 40 students' texts were used to analyze. It should be mentioned that in most cases the entire student's papers in a class have been used and there were no selection involved.

In semester 2 compositions of students attending Advance Writing course were not long enough to be manipulated. Therefore, instead of 40 papers, 90 papers of these students have been used.

After collecting the required data, which text length ranges from 9930 to 10820 words, were all typed and then converted into txt computer files in order to be analyzed in *RANGE*. For each group, *RANGE* has provided number of words in the text, Type/Token Ratio, and the Lexical Frequency Profile. Therefore, we observed how many words of base list 1, base list 2, and base list 3 occurred in each group. There are also some words that do not belong to any of these 3 base lists and they are referred to not in the list figures. These lists were not of concern at this work since presence of high frequency words in a text makes a text richer.

IV. DATA ANALYSIS

Results

Before presenting the data in tables, it seems necessary to give a short definition of the head words in the tables. As mentioned earlier, software *RANGE* provided us with the proportion of frequent and non-frequent vocabulary in students' writings that is the percentage of words used from different vocabulary frequency levels. All the words in a text have been classified into categories of frequency.

This program measures productive vocabulary in the compositions of second language learners. Lexical Frequency Profile or LFP analyzes a piece of writing in terms of the proportion of frequent and non-frequent vocabulary in it, that is the percentage of words used from different vocabulary frequency levels. The entire calculation matches vocabulary frequency lists with the learners' compositions.

The profile is calculated for *tokens* that is all words in the composition, for *types*, that is different words in the composition and for *word families.*

By *Word List* we mean the three lists of most frequent 1000, 2000 and 3000 words by Nation available in RANGE. The third word list forms most of academic or University Words List (UWL).

Range results for semester 2 students

The results of this work have been set out in tables and graphs as follows:

Table 2 presents a thorough description of the three available word lists in terms of word families, tokens and types.

TABLE 2.
TYPES, TOKENS & FAMILIES OF WORDS USED IN STUDENTS' WRITTEN PAPERS IN SEMESTER 2 BY WORD LIST

Word list	Tokens	%	Types	%	Families
One	9092	89.54	777	59.77	487
Two	461	4.54	175	13.46	135
Three	196	1.93	79	6.08	62
Not in the list	405	3.99	269	20.69	????
total	10154	-	1300	-	684

It shows that the written papers of students in semester 2 consist of 684 word families. A word family is the base form of a word that might appear as a headword in a dictionary, plus all the derived and inflected forms of it. Among this 684 word families 487 word families belong to word list one, 135 word families belong to word list two, and 62 word families belong to word list three.

Table 2 also presents a type and token analysis. A token is any occurrence of a word form in the text, regardless of whether it is occurring for the 1st or the nth time. As we can see 9092 of running words are in word list one, which makes up 89.54% of the total words in the text. Then 461 words are in word list two that makes up 4.54% of the running words in total words in the text. Here 196 words belong to word list three that means 1.93% of total words, and 405 words belong to not in the list with 3.99% of the whole words, which are not of interest in this study because they might present words with spelling errors.

With regard to type, any word form which occurs once regardless of how many more times it might occur, table 2 indicates that there are 1300 word forms in students' texts in semester 2. Among this number 777 of word forms belong to word list one that makes up 59.77% of the total. In turn 175 word forms are in word list two with 13.46%, 79 word forms are in word list three making up 6.08%, and 269 word forms do not belong to any of these three lists. This group makes up 20.69% of the total word forms of the texts of the students in semester 2.
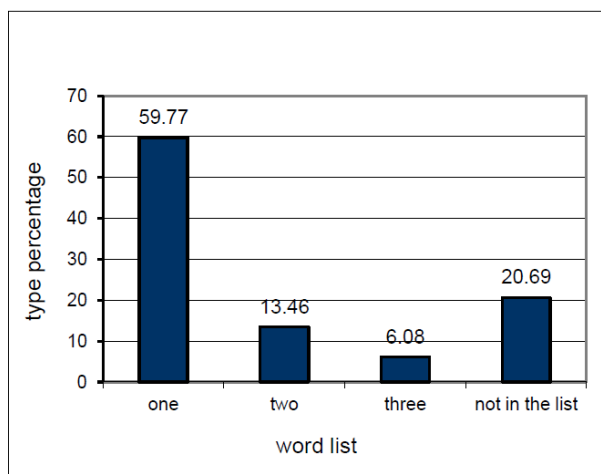
Figure 1. Types of words used in semester 2 students' written papersby word list

Figure 1 graphically shows that students in semester 2 make use of words that belong to word list one adequately. This figure also shows that words from word list two and word list three are not used much by these learners at this stage. It means students are more familiar with the first 1000 most frequent words at early stages than second 1000 high frequency words and word list three. The last one, not in the list, shows the words that do not belong to the available word lists and they are mostly the ones which have spelling errors.

Range results for semester 4 students

Table 3 illustrates that there are 868 word families in students' written papers in semester four. Among this number 555 word families belong to word list one, 215 word families are in word list two, and 98 word families are in word list three.

TABLE 3.
TYPES, TOKENS & FAMILIES OF WORDS USED IN STUDENT'S WRITTEN PAPERS IN SEMESTER 4 BY WORD LIST

Word list	Tokens	%	Types	%	Families
One	8630	86.91	870	57.09	555
Two	707	7.12	287	18.83	215
Three	227	2.29	135	8.86	98
Not in the list	366	3.69	232	15.22	????
total	9930	-	1524	-	868

Table 3 also indicates that 9930 words (tokens) exist in students' texts in semester four. Out of this number 8630 words which makes up 86.91% are in word list one, 707 words that means 7.12% are in word list two, and 227 words which equals to 2.29% are in word list three. And also 366 words do not belong to any of the three word lists which make up 3.69% of the whole text.

Now considering types, there are 1524 word forms at this stage. Among this number 870 word forms or 57.09% belong to word list one, 287 word forms or 18.83% are in word list two, and 135 word forms which makes up 8.86% of the running words belong to word list three. The rest of words that is 232 word forms are in not in the list and makes up 15.22% of the running word forms.
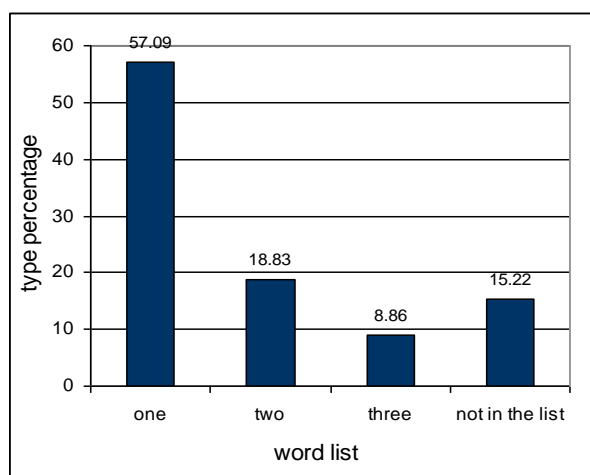

Figure 2. Types of words used in semester 4 students' written papers by word list

As figure 2 depicts we have almost similar results to semester 2, since in semester 4 words in word list one are also used more than other word lists. There are only some insignificant changes in usage of words from word lists two and three.

Range results for semester 6 students

According to table 4 there are 611 word families in written papers of students in semester 6 totally. Out of this number 393 word families belong to word list one and 60 word families are in word list two. The rest that is 158 word families belong to word list three.

TABLE 4.
TYPES, TOKENS & FAMILIES OF WORDS USED IN STUDENT'S WRITTEN PAPERS IN SEMESTER 6 BY WORD LIST

Word list	Tokens	%	Types	%	Families
One	8217	76.56	660	53.10	393
Two	270	2.52	83	6.68	60
Three	1699	15.83	260	20.92	158
Not in the list	547	5.10	240	19.31	????
total	10733	-	1243	-	611

Table 4 also indicates that 10733 tokens are in written papers of students in semester six. There are 8217 words in word list one that makes up 76.56% of running words in the text. At this level we have 270 words from word list two that is 2.52% and 1699 words from word list three that makes up 15.83% of the running words in the text. The rest of words that are 547 or 5.10% are not in any of the three lists.

This table also tells us that there are 1243 types in written papers of students in semester six. It shows that 660 word forms or 53.10% of the word forms are from word list one and 83 word forms or 6.68% of the word forms belong to word list two. It also presents that 260 word forms that makes up 20.92% of the whole text belong to word list three. Again here we have 240 words which make up 19.31% of the running word forms that are not in neither of the previous lists.
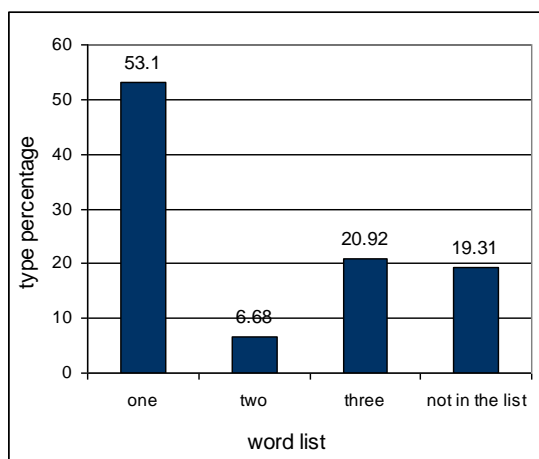

Figure 3. Types of words used in semester 6 students' written papers by word list

As figure 3 shows word forms in word list one are still used in semester six. But there is a significant change in usage of words from word list three which includes academic words.

Range results for semester 8 students

Table 5 provides the descriptive statistics of data selection in semester eight. The results show that there are 478 word families in the written papers of this group. Out of this number 341 word families belong to word list one, 51 word families belong to word list two, and 86 word families are in word list three.

TABLE 5.
WORD LISTS IN SEMESTER 8

Word list	Tokens	%	Types	%	Families
One	8274	76.47	512	59.67	341
Two	384	3.55	73	8.51	51
Three	1410	13.03	134	15.62	86
Not in the list	752	6.95	139	16.20	????
total	10820	-	858	-	478

According to table 5 the sum total of tokens in written papers of students in semester 8 is 10820. As table 5 shows 8274 words are in word list one that makes up 76.47% of the running words, 384 words or 3.55% are in word list two, and 1410 words or 13.03% belong to word list three. The rest is 752 words that do not belong to any list which makes up 6.95% words of the text.

In table 5 the total number of types is presented as 858 word forms. Among this number 512 word forms which makes up 59.67% of the running words are in word list one, 73 words or 8.51% are in word list two, and 134 word forms or 15.62% of the running words belong to word list three. It also provides that there are 139 word forms or 16.20% in the written papers of students at this level that do not belong to any list.

Figure 4 below depicts an overview of types of words used in written papers of students in semester eight.
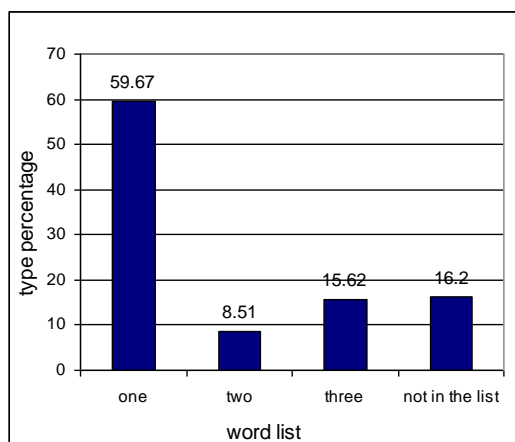

Figure 4. Types of words used in semester 8 students' written papers by word list

It shows that words from word list one is still used at this stage. It also illustrates that words from word list three are used by these students in their writing tasks.

Overall results for semesters 2, 4, 6, and 8

In this section we have an overall view on types of words used by students in semesters 2, 4, 6, and 8 by word lists. Table 6 displays the types of words used in written papers of students in each semester separately according to three available word lists. The last list consists of words that belong to none of the word lists.

TABLE 6.
TYPES OF WORDS USED BY STUDENTS IN SEMESTERS 2, 4, 6, & 8 BASED ON WORD LISTS

Semesters	Types Word list1	Types Word list2	Types Word list3	Types Not in the list
Semester 2	59.77%	13.46%	6.08%	20.69%
Semester 4	57.09%	18.83%	8.86%	15.22%
Semester 6	53.10%	6.68%	20.92%	19.31%
Semester 8	59.67%	8.51%	15.62%	16.20%

According to the table words from word list one has been utilized by each group with almost similar percentage. But word list two does not have the same results. In semesters 2 and 4 words from word list two has been utilized more than in semesters 6 and eight. Table 6 also indicates that words from word list three has been applied by students in semesters 6 and 8 more than those in semesters 2 and four.

Figure 5 graphically shows the way that words from word list one function in each group. Figure 6 illustrates the change in usage of words from word list two among these four semesters. At last, figure 7 displays the trend of growth in words from word list three for semester 2 to semester eight.
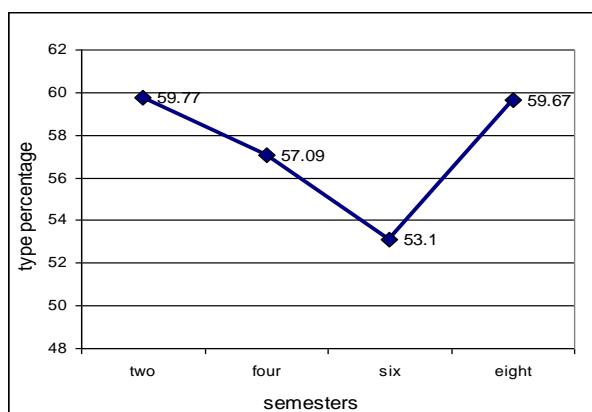

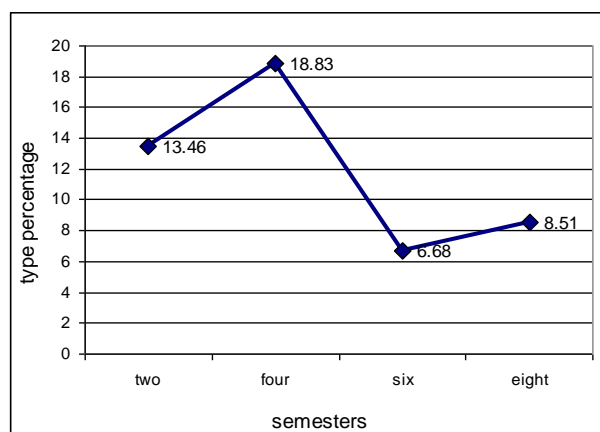Figure 5. Type results of semesters 2, 4, 6, & 8 based on word list 1

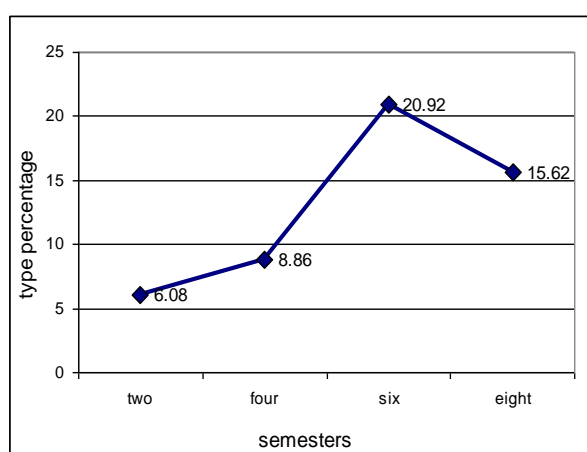Figure 6. Type results of semesters 2, 4, 6, & 8 based on word list 2

Figure 7. Type results of semesters 2, 4, 6, & 8 based on word list 3

A profile for EFL students

Having written text samples of students studying at semesters 2, 4, 6, and 8, we built up a profile for total EFL students studying at semesters two, four, six, and eight. Table 7shows a thorough description of the three available word lists in terms of word families, tokens and types used by EFL students attending this study.

TABLE 7.
OVERALL RESULTS OF WRITTEN TEXT SAMPLES OF EFL STUDENTS BASED ON WORD LISTS

Word list	Tokens	%	Types	%	Families
One	34213	82.17	1433	44.94	732
Two	1822	4.38	481	15.08	316
Three	3532	8.48	438	13.73	247
Not in the list	2070	4.97	837	26.25	????
total	41637	-	3189	-	1295

According to this table there are 41637 tokens in the total samples used by EFL university students. Out of this number 34213 tokens are from word list one, which make up 82.17%, 1822 tokens belong to word list two, which make up 4.38%, 3532 tokens are from word list three, which make up 8.48%, and 2070 tokens do not belong to any list, which make up 4.97% of the running words.

This table also shows that there are 3189 types in the total samples used by EFL university students. Out of this number 1433 types are from word list one, which make up 44.94%, 481 are from word list two, which make up 15.08%, 438 types belong to word list three, which make up 13.73%, and 837 types do not belong to any list, which make up 26.25% of the running words.

This table indicates that there are 1295 word families in the total writing samples of EFL students which 732 word families belong to word list one, 316 word families are from word list two, and 247 word families are from word list three. Since the program works on the 3 frequency levels, it is not able to categorize any words that do not appear in these lists into their word families.
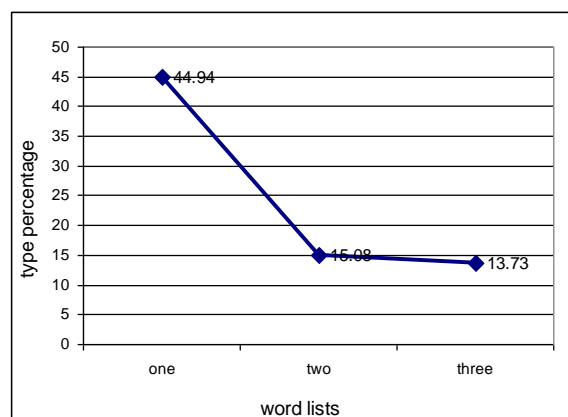
Figure 8. Types results for total written samples in EFL students by word lists

As figure 8 shows EFL learners use the first 1000 most frequent words more than other words based on frequency word lists. It indicates that there is a rapid fall in usage of words from word list two and word list three.

V. DISCUSSION

The results obtained from this analysis showed the frequency levels of the texts of each group. Each figure has compared with available word lists on the program.

Nation (2002) believes words from word lists one and two are necessary for real language use at the beginning of the language courses. On the other hand, words from word list three are useful for reading authentic texts. With regard to word list one, the findings of this work indicated an almost equal status in lexical richness of students in semesters 2, 4, 6, and 8, that is, there is almost the same percentage of words from word list one in their writings. According to Laufer and Nation (1999) the most frequent 1000 words account for around 75% of the running words in formal written texts.

However, at the second word list level, the results showed that there was a significant fall after semesters two and four. In other words at the second 1,000 most frequent words the trend of growth was not significant after semester 4 productively. At the word list three, contrary to the word list two, there was a significant rise after semesters 2 and 4. It showed that there were some degrees of growth in learners' lexical richness after the forth semester.

Laufer and Nation (1995) argued that the texts produced by learners at different levels of proficiency have characteristically different profiles, and that the text profiles change in predictable ways as learners' vocabularies grow. As the results showed in this work the profiles of the students changed at different levels of proficiency and to some extent the changes were predictable. Students in semester 8 who are able to read and understand authentic texts can produce words from word list three as well. Therefore, this view is supported by this study.

To sum up, the evidence presented in the study would indicate that EFL students not only add to their vocabulary knowledge during the years of study, as they are exposed to different sources of English vocabularies, but also use this knowledge in their productive tasks.

A. General Conclusion

A general study of word frequency counts shows that knowledge of the 2,000 most frequent word families constitute a threshold of the words required for basic oral communication (Laufer, 2005). These 2,000 word families make up about 87% of written texts and about 80% of typical academic texts (Nation, 1990).

According to Nation (2002) words from word list one and two are appropriate for real language use at the beginning of the language and words from word list three are useful to read authentic texts. He believes the 2,000 and 3,000 word levels consist the high frequency words that all language learners need to know in order to function effectively in English.

In general, for English a threshold of 3,000 word families is needed for minimal comprehension and 5,000 word families for reading for pleasure (Laufer, 1992 & Schmitt et al., 2001). In order to be successful in academic studies, Nation (1990) believes that it is necessary to be familiar not only with the high frequency words of English, but also with the general academic vocabulary which is common to different academic disciplines. It has been argued that the knowledge of 5,000 word families is required before one can begin to guess the meaning of unknown words in context (Laufer, 1998).

Bearing these notions in mind, this research set out to build up a profile of lexical richness of productive vocabulary from Iranian EFL university students. The results of the study showed an almost significant growth in productive vocabulary of EFL students, though this knowledge is under the suggested threshold (87%) proposed by other studies in the field. Therefore, the results support the hypothesis which claimed EFL university students add to their productive vocabulary knowledge throughout years of study while they are exposed to various English courses.

The evidence presented in the study indicated that EFL students increase their vocabulary knowledge while they are studying different courses and they are able to use this knowledge in their productive tasks. It is obvious that learners

with a large number of words are more competent in a wide scope of language skills than learners with smaller vocabularies, and there is more evidence to be in favor of the view that word skills make an important contribution to almost all aspects of L2 proficiency.

Nation (2002) underlies the critical importance of developing an adequate dependent on the number of vocabularies learners know, especially in the early stages of learning a foreign language with around 3,000 word families being a vital threshold. He believes that a systematic, principled approach to vocabulary development results in better language learning.

Everyone has learnt that the more often a learner encounters a vocabulary item the higher the chances for him to acquire the word. Consequently exposure of learners to rich and varied input is crucial.

According to the results of this study, students attending semesters 6 and 8 possess and use words from word list three more than students in semesters 2 and 4. It indicates that students add to their academic vocabulary knowledge through various input which are available for them each semester. One of these resources is the books that are presented for every individual course. It has been argued that most if not all vocabulary development (in L1 & L2) takes place as learners attempt to comprehend written input (Krashen, 1989, 1993).The increase in productive vocabulary knowledge of students as they get to higher levels of education reveals their ability to control and use of receptive items which is shaped by acquiring a complete set of information of those words. This mastery is obtained by a number of repetitions, practices, and exposures of the lexical items. However, language teachers and program developers should pay more attention to the ways of promoting vocabulary more effectively among language learners. This study suggests that the high frequency words are worth special attention and thus should be taught directly. The findings suggest a more prominent position for vocabulary instruction in foreign language teaching programs.

B. Pedagogical Implications

The findings of this study can also be useful in the field of language pedagogy. Such findings can be used by language teachers and administrators in a pedagogical context to make decisions concerning whether or not learners have enough lexical resources necessary to function effectively in the target language.

According to Meara (2005) once it is understood how language learners learn language through the different mediums by which language is received or created, then teachers will be better able to provide timely and appropriate interventions. The evidence presented in this study indicated that EFL students expand their vocabulary knowledge when they get to higher levels of proficiency and they are also able to control and use this knowledge in their productive tasks. This ability indicates that the learners take advantage of the input available for them throughout different courses at university.

With regard to resources available for learners at university, they contain different kinds of lexical items which vary each semester depending on the subject of the course without considering the level of proficiency of the learners. At higher stages of education the content of these resources seems to be more important than the language itself, so the academic needs which demand rich and varied input have been ignored.

To improve and develop learner's vocabulary knowledge, teachers and language programmers should pay more attention to both the content and the vocabulary needs of learners when they select related resources in higher levels of education. These resources need to have the necessary lexical items for the learners which include high frequency words, low frequency words, and academic words.

Hazenberg and Hulstijn (1996) calculated that foreign students reading university texts need to have 10,000 word families at their disposal. At higher levels of education in addition to explicit learning, a learner should learn vocabulary autonomously. Learners can expand their vocabulary through reading, combined with the development of a raised awareness of vocabulary learning strategies. And it is the teacher who can create this awareness among students by developing appropriate tasks for them. Therefore, the teachers should be conscious of the feedback they need to have from the learners first. Then they need a pedagogical plan for this purpose which demands systematic presentations of lexical items to students in each semester.

Nation (2005) believes teaching can effectively deal with only a small amount of information about a word at a time. The more complex the information is, the more likely the learners are to misinterpret it. We need to see learning any particular words being a cumulative process where knowledge is built up over a serious of varied meetings with the word. At best, teaching can provide only one or two of these meetings. The others involve deliberate study, meeting through meaning focused input and output, and fluency development activities.

As discussed earlier,if vocabulary is accepted as an essential building block of language, then it naturally needs to be assessed in some way. Meara (2005) believes instead of trying to practice receptive items as much as possible, which may only serve to reinforce existing links, a more effective strategy would be to stress associations from already (productive) known words to new or receptive items. The teachers can involve students with vocabulary learning by different methods, such as quiz, composition, and discussion, interactively. Teachers should increase reading and writing tasks and draw their attention to the importance of vocabulary knowledge in these tasks.

Lexical knowledge and lexical use take different paths of development. Therefore, the selection of resources for each semester should be investigated thoroughly which includes the classification of lexical items according to their levels of frequency and the focus on both content and language skills. The distinction between high and low frequency words is

important for teachers because it makes it necessary for teachers to know what stage their learners are at in their vocabulary development.

REFERENCES

[1]   Anderson, R. C., &Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews*. Newmark, DE: International Reading Association.
[2]   Baayen, H. (2001). Word frequency: Distributions.Dordrecht, Kluwer Academic publishers.
[3]   Carter, R. (1998). Vocabulary: Applied linguistic perspectives (2ⁿᵈed.). London: Routledge.
[4]   Cobb, T., &Horst,M. (2002). Is there room for an academic word list in French? In B. Harley and J. Hulstijn (Eds.), *Vocabulary in a second language: Selection, acquisition and testing, processing of Haifa-Leiden conference on vocabulary*. Amesterdam: Benjamins.
[5]   Hazenberg, S., &Hulstijn, J. (1996). Defining a minimal receptive second language vocabulary for non-native university students: an empirical investigation. *Applied linguistics, 17*(2), 145-163.
[6]   Herdina, P., & Jessner, V. (2002). *A dynamic model of Multilingualism: Changing the psycholinguistic Perspective*. Clevedon: Multilingual Matters.
[7]   Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the input hypothesis. *The modern language journal, 73*,404-440.
[8]   Krashen, S. (1993). *The power of reading.* Englewood, CO: Libraries Unlimited, Inc.
[9]   Laufer, B. (1992). How much lexis is necessary for reading comprehension? In Arnaud, P., and Bejoint, H. (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.
[10] Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*(2), 255-271.
[11] Laufer, B. (2005). Lexical Frequency Profiles: From Monte Carlo to the Real World. A Research to Meara (2005).University of Haifa.*Applied Linguistics, 26*(4), 582-588.
[12] Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics, 16,*307-22.
[13] Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language testing, 16*(1), 33-51.
[14] Meara, P. (2005). Lexical Frequency Profiles: A Monte Carlo Analysis. *Applied Linguistics, 26* (1), 32-47.
[15] Meara, P., Lightbown, P. M., & Halter,R. (1997). 'Classroom as lexical environment'. *Language Teaching Research, 1,* 28-47.
[16] Muncie, J. (2002). 'Process writing and vocabulary development: Comparing lexical frequency profiles across drafts'. *System, 30* (2), 225-35.
[17] Nation, I. S. P. (1990). Teaching and learning vocabulary. New York: Newbury House.
[18] Nation, I. S. P. (2001). Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.
[19] Nation, I. S. P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press.
[20] Nation, P. (2005). Teaching Vocabulary.*The Asian EFL Journal, 7*(3), 47- 54
[21] Nation, P., &Heatley, A. (1996). VocabProfile, word and range: Programs for processing text. LALS, Victoria University of Wellington.
[22] Read, J. (2000). Assessing vocabulary.Cambridge: Cambridge University Press.
[23] Reid, J. M. (1993). Teaching ESL writing. Englewood Cliffs, NJ: Regents/Prentice Hall.
[24] Richards, J. C. &Renandya, W. A. (2002). Methodology in language teaching. Cambridge: Cambridge University Press.
[25] Schmitt, N., Schmitt, D., &Clapham, C. (2001). Developing and exploring the behavior of two new versions of the vocabulary levels test. *Language testing, 18,* 55-88.
[26] Steel, L. (1995). 'A self-organizing spatial vocabulary'. *Artificial life Journal, 2*(3), 319-32.
[27] Steel, L. (1996). 'Emergent adaptive lexicons' in P. Maes, (ed.), *From animals to animates 4: Processing of the forth international conference on simulating adaptive behavior* (pp. 562-7). Cambridge MA: MIT Press.
[28] Steel, L., &F. Kaplan. (2002). Bootstrapping grounded word semantics in T. Briscoe (ed.), *linguistic evolution through language acquisition: Formal and computational models* (pp. 53-73). Cambridge: Cambridge University Press.

Nazli Azodi received her B.A. degree in TEFL at Najafabad Azad University, Isfahan, Iran in 1996. Then she received her M.A. in TEFL at Khorasgan Azad University, Isfahan, Iran in 2007. She is a Ph.D. student in TEFL at Khorasgan Azad University, Isfahan. She has been teaching English to Iranian EFL learners for more than 10 years in several English Institutes and she, currently, is teaching at both Khorasgan Azad University and University of Applied Sciences and Technology in Isfahan. She has also contributed to the International Journal of Research Studies in Educational Technology. She is interested in second language acquisition, language teaching methodology, phonology, and methods of improving the productive vocabulary.

Fatemeh Karimi, born in Rasht, Iran, she is an English instructor at Islamic Azad university khorasgan branch. She received her M.A. degree in TEFL from Tarbiat Moallem University of Tabriz in 2006. Her research interests are language testing and research.

Ramin Vaezi, born in Esfahan, Iran, he is a Ph.D. student in Teaching English as a Foreign Language (TEFL) at Islamic Azad University Khorasgan (Isfahan) Branch in Iran. He has also co-authored a book on English reading comprehension. He has presented and published articles in different national and international conferences and journals, such as the Second Regional Conference on English Literature and Applied Linguistics, Journal of language Teaching and Research, Cross-Cultural Communication Journal, etc. He has also reviewed a number of articles for journal of Languages and Culture (JLC). Moreover, he is now an editorial board member of Open Applied Linguistics journal. Mr. Vaezi has been teaching English in different universities, institutes of higher education, and language schools in Tehran and Esfahan for the last nine years. He is an English teacher and university instructor now.