Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.4, No.6, 2013
www.iiste.org

Lexical Richness, a reliable measure of Intermediate L2 Learners' current status of acquisition of English language

1. Sayed Kazim Shah
Department of English Linguistics, Government College University, Faisalabad, Pakistan
E-mail: kazimkazmi@gmail.com

2. Ayesha Asghar Gill
M.phil Scholar, Department of English Linguistics, Government College University, Faisalabad, Pakistan

3. Dr. Rashid Mahmood
Department of English Linguistics, Government College University, Faisalabad, Pakistan
E-mail: ch.raashidmahmood@gmail.com

4. Muhammad Bilal
Department of English Linguistics, Government College University, Faisalabad, Pakistan

Abstract
This article aims to explore the utility of the relationship between lexical richness and size as an indicator of acquisition status of English language of L2 learners of intermediate level, having rural background on the basis of their self -written output. 126 students' essays were used to measure the lexical richness (126 students of Sem-I and 63 students of Sem-II) Lexical Frequency Profile was used to sort it out. Its values discriminated students of different proficiency level and displayed L2 Learners vocabulary size in use. LFP result's consistency and legitimacy was obtained by comparing its result with an independent and separate measure of vocabulary size, VLT. The result showed that lexical richness has a direct link with vocabulary size (receptive vocabulary) of L2 learners. It discusses the utility of the inference based on the lexical richness of L2's written text for monitoring purpose of language acquisition process of L2 learners and to determine appropriate strategies for the desired growth of vocabulary size.

1.1 Introduction

Vocabulary size is an important factor for, mastering L2 language (Schmitt, 2008). It is essential for meaningful communication and effective reading comprehension (McCarthy, 1990). In Pakistan, most of the educational material, taught and used, basically is written by and for the natives. It is creating a hindrance to grasp the conceptual knowledge of English language for L2 Learners, especially less proficient ones .In order to increase their comprehension and productivity, in both spoken and written language skills, we need to know the present status of L2 learners as a reference point (Nation, 2001). In this regard, the measure of lexical richness of L2 learners facilitates to quantify the desired level of vocabulary level at any threshold of education. It will define the requirement of vocabulary level to perform different tasks of academic education. Such statistical information determines the quality of factors that affect the quality of language learning process and on the other hand it gives clear cut relationship between vocabulary knowledge and use on the basis of written text.

1.2 Literature History

Vocabulary learning is a focal point of second language acquisition (Gao, 2003). It means that vocabulary gives a clue about the different areas of language based on different aspects of vocabulary such as size, depth, receptive and productive level on one hand, the way vocabulary is taught and influencing factor such as educational background and the environment of learning on the other hand. This study focuses to sort out a reference point with the help of lexical richness that can be served as measuring unit for the induction of students in a particular level for the positive development of vocabulary growth.

One of the general issues that emerge in relation to vocabulary is the distinction between productive and receptive vocabulary. This study assumes the R/P bipolar rather than binary distinction between these two aspects of vocabulary (Melka, 1997). The distance between these two points is thought as the increased familiarity of a word on a continuum. That shows a gradual movement along with the continuum with the increase of familiarity with word from recognition form, towards recognition meaning, recall form and recall meaning (Laufer & Goldstein. 2004). Similarly, there is always a problem while measuring productive vocabulary accurately of the respondent as compared to recessive vocabulary. The basic reason is that, productive vocabulary is always subjected to context. It calls accuracy of the measurement of productive vocabulary in question (Meara &Fitzpatrick, 2000) because such association gives inconsistent results.

Many past researches have reported a gap between receptive and productive vocabulary. Receptive vocabulary level is greater than productive vocabulary knowledge. This disparity is resolved by following the Reads

conceptual distinction for receptive and productive vocabulary knowledge. He proposes that this confusion is to look as receptive vocabulary knowledge and productive vocabulary knowledge as separately by defining the former as recognition and later as recall or as understanding and use. If one aims to calculate recognition and recall vocabulary then discrete tests are best option. Similarly, if one wants to calculate understanding and use then embedded tests are more suitable (2000). As the study aims to find out the vocabulary understanding and use, so L2 learners' written compositions were used to get estimation. Moreover, the context of learning, rural areas background has been precisely defined. So, the findings can be compared with other research works as well.

In past works, Productive measures were usually in the form of tests. Respondents were required to complete the given word, provided with first two or three letters clues in a given context as PVL productive level test (Laufer & Nation, 1999). Likewise there were such measures where respondent was required to write as many responses as possible to a given stimulus such as Lex 30 (Meara & Fitzpatrick, 2000). These responses were analyzed with the help of most frequent used words lists. Their responses might lie in 1st 1000 most frequent used words, or in 2nd 1000 most frequent words but if the responses lied beyond 2000 category, less frequent words, the productive vocabulary was considered as larger one.(Laufer,1998; Laufer&Nation,1999; Akabarian,2001; Meara& Fitzpatrick,2000; Morris & Cobb,2004; Muncie,2002;Milton,2008;2009; Read,2000). As a resultant, this study has used students' essay as a specimen to measure vocabulary size.

It means that responses, to a stimulus and filling a word in a provided space according to the context, are different from the production of words in a composition. It requires different measure to gauge its strength. Vocabulary used in composition may be measured in number of ways. Laufer and Nation have reported four of the most prevalent measures. First is Lexical originality, it is calculated by adding up the number of unique words associated to one writer in the group divided by total number of token used. This method provides information about an individual vocabulary use in relation to his /her fellow beings. Such statistics is useful but it cannot stand alone as it varies across different compositions of an individual or it would change as the group changes. Second is Lexical Density, it is the percentage of lexical word present in a text divided by the total number of words. Lexical words stands for noun, adjectives, verb and adverbs. The short coming of this method is, that it ignores syntactic structure of composition and other cohesive devices present in a given composition. Third is Lexical Sophistication, it is obtained by comparing the advanced level words used in the composition with the total number of the words used in it. This method has a drawback. It is based on the number of advanced words, which are relative to the exposure and learning. It means this is not a reliable measure of productive vocabulary. Fourth is the Lexical variation, it is measured by type/token ratio. If a composition has greater number of different words, it would be a better composition. This method is only sensitive for different numbers of words but it does not point out the quality of different words used in composition (Laufer & Nation, 2001).

Likewise, there is another kind of measurement, Lexical Quality formula (East, 2009), it is the sum of numbers of types and rare words minus twice the number of lexical errors. This method has some ambiguity: the concept of rare words and errors are not properly defined. The errors found in Learner's composition such as misspellings, derivational errors, semantic errors and inappropriate word choice (Read, 2000). This loop hole suggests that results have subjective nature of judgments and questions the checker's reliability with reference to relative seriousness, for different error types. Furthermore, for the sake of lexical richness, learner's lexical knowledge and vocabulary size is measured without subtracting them from the existing vocabulary as the LQ formula does.

Similarly, T-unit length and error free T-unit length are occasionally used as an indicator of lexical richness measure. It has many problems such as this measure relies on syntactic property of writing and the length of T-unit: based on verbosity than lexical richness. In addition, sometimes in low proficiency learners, main and sub ordinate clauses are difficult to identify. Overall the calculation of T-unit length is based on subjective interpretation of the reader. In the consequences of above mentioned draw backs, this study uses Lexical Frequency Profile for the measurement of lexical richness of written production.

LFP shows the relative proportion of words from different frequency levels. This study aims to verify the utility of lexical richness as an indicator of vocabulary size. It has used this feature of LFP to differentiate low and high vocabulary level students, so it proposes two LFP measures: one for less proficient students, the other is for advanced ones. The main reference for measuring lexical richness for less proficient students is between the first 1000 most frequent words, the second 1000 and any other vocabulary and for advanced learner the reference for measuring LFP is second 1000 most frequent words, the academic vocabulary (UWL) and words not in any lists would organize the count of LFP. It will be calculated as follows, suppose a learner's composition has 200 word families. Among the 200, 150 belong to first 1000 most frequent words, 20 to the second 1000, 20 to the UWL and 10 are not in the list. To calculate LFP, firstly these numbers are converted into percentage out of the total of 200 word families. The LFP of the composition is therefore 75%, 10%, 10 %, 5%. LFP uses word family as a

counting unit for the measurement of lexical richness of a composition. It closely matches with learners' view of words. The learners at high proficiency level would not find any difficulty in seeing that happy, happiness, happiness, happiest, happily and unhappy are closely related. Bauer and Nation (1993) provided a guideline for defining the member of a word family. It is used as a reference point in this study.

The entire process is done by a computer programme Range. It has vocabulary lists of frequent words. The program uses the text in the form of ASCII format. It compares the given text against the vocabulary lists based on frequency of words. It marks the words in the given text and enlists the words from the text in types and families according to the list they occurred in.

LFP has several benefits in comparison to the above said measurements of lexical richness. It is better than lexical originality as it is independent of learner's reference group's lexical richness. In addition, it does not change with the change of reference group and it does not consider syntax and text cohesiveness.

Since, Lexical Frequency Profile shows advanced vocabulary like lexical Sophistication , but unlike Lexical Sophistication LFP gives more detailed account of different types of words and LFP is based on word frequency rather than in terms of lexical syllabus or words from given reading material. Lexical sophistication depends on discipline of lexical environment, whereas, LFP can compare texts of group related different educational systems or groups learning language outside the classroom. Similarly, LFP has an edge on Lexical variation, as it can differentiate between subjects having frequent and less frequent vocabulary.

In a nut shell LFP is more reliable than any other lexical richness measure as it is independent of subjective decisions based on topic, sub-topic, elaboration and thematic unit. If a word is not used correctly than it would not be considered as a part of learner's lexicon.

2. Study

This research aims at establishing the reliability and validity of LFP to discriminate the status of Language of L2 learners of intermediate level, belonging to rural background, on the basis of lexical richness measured from free written composition. It is based on the following research questions:

Q 1: Does LFP able to discriminate between students' level of language acquisition on the basis of lexical richness?

Q 2: Does LFP result correlate with another independent measure as VLT?

Q 3: Does Lexical Frequency Profile change across two pieces of writing?

2.1 Subject

The subjects were L2 learners of intermediate level. They belong to different rural areas of Pakistan. There were three groups of learners. First group was semester 1 students and second group was semester 2nd students and third group was semester 3rd students. They were native speakers of Punjabi. There were 63 students in each group. They were 189 students altogether.

3. Method

3.1 Data Collection

Two compositions were written by respondents as a take home assignment. They were required to write 250-300 words. The topics of the composition were of general nature. None of them needed expert knowledge of particular subject matter. Moreover, reliability of data was ensured by keeping a check on not using any help either from tutors or any other resources. The topics for group 1 were same for all. Likewise, group 2 had same topics but different from group 1. For the sake of interest of students a motivation was given to all students that these compositions would be a part of their assignment which would be included in final semester assessment of grades.

Test B

The respondents were given the active version of the Vocabulary Level Test to elicit their vocabulary knowledge. It had 5 frequency levels in given sentences (the second 1,000 words, the third 1000, the fifth 1000, the University Word List and the tenth 1000). According to Nation (1990), the 2000 and 3000 word levels contain the high frequency words that all learners need to know in order to function effectively in English. The 5000 word level represents the upper limit of general high frequency vocabulary that is worth spending time on in class (Nation, 1983). Words at the academic level should help students in reading academic text books and 10,000 word levels covers the more common lower frequency words in the language. This level is not used as this level is out of the reach of the learners.

In this version of the VLT, there were 10 clusters at each level. Each frequency level of the test comprised 6 words and 3 definitions. Testees were required to match target words with their corresponding definition.

3.2 Data Processing

All the compositions were entered into the computer in the form of ASCII files. A wrong derivative of a word was not considered an error. Proper nouns were deleted from the samples. The following list was produced for every composition: first thousand most frequent words, second thousand most frequent words, university Word

List, and not-in-the-lists word list. For each composition LFP was calculated, on the basis of the proportion of word families at each of these four levels. it was computed by a free software Range.

3.3 Results and Discussion

Table3.1: Mean percentages and standard deviation of word families at different frequency levels

	1st 1000		2nd 1000		AWL	
	Comp 1	**Comp 2**	**Comp 1**	**Comp 2**	**Comp 1**	**Comp 2**
Group 1	71.0097	73.1406	14.3797	13.1229	14.6062	14.0156
SD	4.35	16.87338	3.2026	2.89385	2.45217	2.377
Group 2	62.839	73.0832	17.6071	12.4529	14.4400	14.4174
SD	3.89	16.88262	1.928	2.46190	2.73875	2.727
Group 3	70.8891	73.0247	17.2921	17.8321	11.7912	5.4550
SD	5.74	4.567	2.50	5.22840	3.45083	2.61504
F-test	40.910	0.001	15.957	20.983	10.009	131.174
P value	.000	0.999	0.000	0.000	0.000	0.000

First of all the issue of validity was focused and the research questions related to it were justified

Q 1: Does LFP able to discriminate between students' level of language acquisition on the basis of lexical richness?

Q 3: Does Lexical Frequency Profile change across two pieces of writing?

Table: 3.1 presents the mean percentages of words used by the three groups of learners at different frequency levels. It portrays the results of an ANOVA (comparing of means of the groups at each frequency level) and F test degree of the significance. The means were considered different from each other when the p value was not higher than 0.05.

Table: 3.1 showed that three proficiency groups of learners were significantly different from each other in the 1st 1000 most frequent word families percentage. In a post hoc analysis of the ANOVA by using the Duncan procedure the following differences were observed. In composition1, the Group1 was different from Group 2 &3. The Group1 subjects used considerable more words of highest frequency as compared to Group2 and Group3 respectively. In composition 2 relatively the usage of most frequent words remained same, as it was the literary essay. As a whole, the approach and vocabulary had minor differences. The three groups were not significantly different from one another in 2nd composition of 1st 1000 words.

In second 1000 words group percentage Group 1 is different from other two groups. Likewise, Group 2&3 were not significantly different from one another in composition 1. When we analyzed composition 2 of second 1000 group there was a difference in all Groups words usage. Group 1 &2 used less UWL group but still used more UWL words than Group 3. Same trend was present in Comp 2 for UWL group. In addition, Group 3 used fewer words of UWL groups as compared to Group 1&2. Above all, in all compositions there were no Not-in-List group words. This trend shows that less proficient students make use of the first 1,000 most frequent words; there was a trend for 2nd 1000 group usage among them. But there was fewer word usage of sophisticated vocabulary, the UWL group, among all students. These differences fall in line with the concept of language proficiency which undertakes that proficient user's uses rich vocabulary. LFP has traced out this difference clearly. This result fortifies its validity.

Research question 'c' dealt with concurrent validity of the LFP measure. For this purpose the best available source is vocabulary Level Test as it has been approved to show satisfactory result on an implicational scaling of the levels.

The Vocabulary Level Test was correlated with each component of the LFP. It was required to use composite grades of VLT because they have no levels corresponding to the first 1000 and 'not-in-the-list'. The correlations are presented in Table: 2 as it is evident from the analysis

Table: 3.2 Correlations between the LFP and the Vocabulary Level Test

	1st 1000		2nd 1000		UWL	
	Comp 1	Comp 2	Comp 1	Comp 2	Comp 1	Comp 2
VLT/LFP	0.104	-0.136	0.125	0.250*	-0.207*	-0.341**
P value	0.299	0.174	0.211	0.011	0.037	0.000

The learners VLT grades had no significant correlation with 1st 1000 and 2nd 1000 most frequent words level. They have significant correlation with UWL as the students, more proficient in English, had used more high frequency words in their composition (Engber, 1995). It could be interpreted that as 1st 1000 level includes almost all function words and most basic lexical words. Since, these are prerequisite for written expression.

That's why L2 learners only try to learn these words in order to keep the ball running. These words are sufficient enough to get through the exams. They did not feel any utility in developing their language. This is a attitude is in accordance with Laufer's (1991) "active vocabulary threshold hypothesis". It says, when learners do not perceive the need to enrich their active vocabulary any more, their effort for development of vocabulary will stop.

Accordingly, for the reliability of LFP, a within-subject analysis was carried out on the two sets of compositions written by each student. Since, there was no words fell in not-in-the-list level; it was not included in this comparison. Table: 3 presents the values of the various Comparisons: matched t-tests for the individual levels.

It can be seen from Table 3.3, group 1, 3 exhibited same non-significant profiles in the two compositions at 1st 1000 and 2nd 1000 most frequent words level. At the same time, group 2 had significant difference at UWL level. These results suggest that LFP is stable besides advanced learners, whose vocabulary becomes unpredictable across different samples of writing. This is also supported by Laufer's(2005) contention that lexical knowledge and lexical use develop along different paths

Table3. 3: Comparison of LFP in the two compositions

	1st 1000 t-paired	2nd 1000 t-paired	UWL t-paired
Group 1	-0.785	1.876	1.086
	P=0.438	P=0.69	0.286
Group 2	-3.625	9.113	1.000
	P=0.001	0.000	P=0.325
Group 3	-1.598	-0.671	9.998
	P=0.120	P=0.507	P=0.000

This analysis suggests another approach towards its interpretation as first 1,000 words are mostly function words and lexical words, they are basic words for written expression but real worth of piece of writing is determined by more advanced frequency levels (Santos, 1998). That's why group 1 &2 did not show significant differences. It exposes the learning style of these L2 Pakistan intermediate learners, they commit vocabulary to memory but they are unable to bring it in their linguistic competence. Their vocabulary remained stable as their learning is majorly classroom based activities, such as, listening to the teacher , taking notes and memorization.

When a comparison was made between the results obtained in Table 2 and Table 3, none of the differences between the two essays is significant, showing that this profile is a better measure of lexical richness in general, and specifically it is better for the advanced students in our sample.

Conclusion:

In short, this study suggests that lexical richness is a valid and reliable measure of tracking the process of language acquisition from the written composition of the L2 learners. It is sufficient enough to rely upon as it correlates well with another independent measure, VLT. It is easily managed due to computer software based analysis except for the preparation of written text. This study has used word family as a counting unit; it exposes, L2 learners have weak concept of word classes of a given word. It means that word knowledge is an important factor for the L2 vocabulary acquisition process. This deficiency contributes a negative effect on language acquisition. Since, they have low vocabulary at their hand for use. It suggests teachers to plan the language acquisition programme according to the needs of the learners and advocates, developing a critical approach according to the needs of learners, for their teaching style and syllabus for a manifold output.

It provides a detailed account of vocabulary richness report of a respondent. It shows that, as in past, vocabulary above elementary level was considered as mostly incidental in nature, is not a reality. Likewise, this analysis has the capability to be used as an investigative measure for the output of teaching method. Pakistani learners having rural background are developing a basic competence in language use; they need teaching of vocabulary for the acquisition of English language in a well-planned manner, on regular basis (Moras et al. 2001).

This study put forward that, vocabulary learning strategies should be the indispensable part of syllabus. Teachers should adopt various techniques such as auditory encoding, semantic encoding, contextual encoding and self-initiation to promote the use of unfamiliar words by students for proper language development.

The result of lexical richness is a useful analysis for teachers to determine the kind of attention they should give to develop vocabulary for particular groups of learners. It means that teachers can identify and remedy the deficiency in their students' vocabulary with a targeted aim.

Lexical richness is reliable information for planning successful syllabus according to the needs of learners and for planning and judging the syllabus. Moreover, range, computer software for the assessment of lexical

richness, is free of cost and time saving instrument. It is the cheapest way to dig out the foundation of effective Language learning process especially in underdeveloped countries scenario like Pakistan. It can easily specify the weak area and which vocabulary should be taught (Laufer& Nation, 1999).

This study suggests, lexical richness might be used as a placement criterion for grouping learners into different levels within an academic programme. It will help to address their problems more successfully.

References:

1. Akbarian, I. (2010). The relationship between size and depth for ESP/EAP learners. *Pre-published online in System x(x).* , 1-11.
2. East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing* 14 (2).88-115.
3. Engber, C. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing* 4 (2).138-155.
4. Laufer, B. and I., S., P. Nation. 1999. "A vocabulary size test of controlled productive ability". *Language Testing* 16 (1).33-51
5. Moras,S., Carlos,S.,(2001). Teaching vocabulary to advance students. A Lexical Approach. Karen's Linguistics Issue online Journal.Retrieved from www3.telus.net/linguisticissues/teachingvocabulary.html.
6. Meara, P. M., & Fitzpatrick, T. (2000). "Lex30: an improved method of assessing productive vocabulary in an L2". *System* 28 (1). 19-30.
7. Morris, L.,& Cobb,T. (2004). Vocabulary profiles as predictors of the academic performance of TESL trainees. *System 32(1)* , 75-88.
8. Muncie, J. (2002). Process writing and vocabulary development: comparing lexical frequency profiles across drafts. *System* 30 (2).225-235.
9. Milton, J. (2008). French vocabulary breadth among learners in the British school and university system: comparing knowledge over time. *Journal of French Language Studies* 18 (3).333-348.
10. Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition.* Bristol: Multilingual Matters.
11. Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge: Cambridge University Press.
12. Nation, I. S. P., & Beglar, D. (2007). "A vocabulary size test". *The Language Teacher* 31 (1).9-13.
13. Read, J. (2000). *Assessing Vocabulary.* Cambridge: Cambridge University Press.
14. Santos, T. (1988). Professors' reactions to the academic writing of nonnativespeaking students. *TESOL Quarterly 22 (1)* , 69–90.
15. Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography* 6 (4). 253-279.

This academic article was published by The International Institute for Science, Technology and Education (IISTE). The IISTE is a pioneer in the Open Access Publishing service based in the U.S. and Europe. The aim of the institute is Accelerating Global Knowledge Sharing.

More information about the publisher can be found in the IISTE's homepage: http://www.iiste.org

CALL FOR PAPERS

The IISTE is currently hosting more than 30 peer-reviewed academic journals and collaborating with academic institutions around the world. There's no deadline for submission. **Prospective authors of IISTE journals can find the submission instruction on the following page:** http://www.iiste.org/Journals/

The IISTE editorial team promises to the review and publish all the qualified submissions in a **fast** manner. All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Printed version of the journals is also available upon request of readers and authors.

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar