

LEXICAL DEVELOPMENT ACROSS SECOND LANGUAGE PROFICIENCY LEVELS IN INDIAN CONTEXT: AN EXPLORATORY STUDY

T. S. Safwan. Hudawi

Research scholar (Ph.D)

School of English Language Education

The English and Foreign Languages University,
Hyderabad, Telangana, India

Abstract

The present study was investigated how second language (L2) productive vocabulary develops over time by targeting multiple dimensions of word knowledge. The research question addressed were a) Is there a difference in the rate of lexical variety between learners of different proficiency level? b) Is there a difference in the rate of lexical density between learners of different proficiency level? c) Which kind of lexis features more in both the proficiency level?

The study drew on a 2071- word written corpus produced by 10 intermediate and secondary level learners from Hyderabad regional school, their L2 is English. They were 13 year old (8th grade) and 15 years old (10th grade) English learners. Data analyses were based on type token ratio formulated by Guiraud, 1957 was applied for finding the lexical variation between two groups, and Ure's method was used to find out lexical density between two groups. Findings indicate that learners' written vocabulary did not have significant difference in their rate of lexical variety and density according to their difference in proficiency level. Contradictory to the researcher's hypotheses 8th grade pupils were more proficient than 10th grade pupils. 10th grade pupils outperformed 8th grade in using nouns, adjectives and adverbs except in verb which shows a less difference regarding their proficiency level.

Keywords: lexical variety, lexical density, proficiency, vocabulary, second language acquisition

1. 0. INTRODUCTION

The fundamental role that lexis plays in second language learning and teaching has been repeatedly acknowledged in theoretical and empirical Second Language Acquisition (SLA) vocabulary research. Singleton, in the introduction to his seminal book on the mental lexicon, asserts that “the major challenge of learning and using a language – whether as L1 or L2 lies not in the area of broad syntactic principles but in the ‘nitty-gritty’ of the lexicon” (1999:4), an idea also shared by Hunt and Beglar (2002:2) who argue that “the single most important task facing language learners is acquiring a sufficient large vocabulary (Levis, 2008:8), or that “the most striking difference between foreign learners and native speakers is in the quantity of words each group possesses” (Laufer, 1998:225). Lexical variation (also known as lexical diversity) is measured by looking at the range of different words used across a text, or in other terms, the

extent to which repetition is lacking (Malvern, Richards, Chipere & Durán, 2004; McCarthy & Jarvis, 2013) and Lexical density is measured by the proportion of content words (nouns, lexical verbs, adjectives and non-grammaticalised adverbs) used in a text (Schmitt, 2000). The measurement for both variables is briefly prescribed in the succeeding paragraphs.

The relevance of the lexical dimension of SLA justifies the explosion of theoretical and empirical research in the area, particularly since the 1990's (cf. Bongaards & Laufer, 2004; Carter, 1987; de Groot & Van Hell, 2005; McCarthy, 1990; Cenoz, Jessener & Hufeison, 2003; Coady & Hackin, 1997; Lengyel & Nauracsics, 2007; Meara, 2002; Nation, 2001; Schmitt & McCarthy, 1998; Singleton, 1999). Collectively, this research has shed light on various themes in the domains of theory and pedagogy.

From the perspective of theory, a central focus of scholarly discussions and empirical investigations has been the very concept of vocabulary knowledge and lexical competence (Pavlenko, 1999; Nation, 2005; Singleton, 1999), together with the related questions of whether or not the unit of analysis teaching should be the "word" (Gardner, 2007, Hunt & Beglar, 2005) or the question about the dimensions of breadth and depth in lexical competence (Laufer & Goldstein, 2004; Nation, 2005; Weshche & Parabut, 1996). Another area of contention relates to the nature of the mental lexicons particularly regarding the debate as to whether or not the languages in the bi/ multilingual mind are represented separately (cf. Cook, 1992; Dijkstra, 2003; Kroll & de Groot, 1997; Navracsiscs, 2007, Singleton, 1999, 2007; Smith, 1997; Schreuder & Welten, 1993), the lexicon's internal structure, and its accessibility in comprehension and production (Altarriba, 2006; Chapelle, 1998, de Groot, 2002; Meara, 2005). Considerable efforts have also gone into the empirical study of vocabulary strategies in terms of which strategies are employed by L2 learners in their attempt to acquire and use their L2 vocabulary, the effectiveness of their deployment, and the variables influencing their use (Nation, 2005; Nyikos & Fan). Closely related to this last issue is the study of the role played in vocabulary acquisition representation by individual differences such as gender, age, and proficiency related differences (cf. Cenoz, 2002; Ferre, Sanchez- Casas, & Guash, 2006; Grace, 2000; Gu, 2002; Henning, 2003; Schmitt, 1997).

The present paper explores quantitative aspects of the development of lexical richness in English as a second language specially focusing on the acquisition of lexical diversity (saying how many different words are used), and lexical density (saying what is the proportion of content words in the text) of adolescents in Indian context. Here we focus on the development of the lexicon in written texts in English of pupils in 8th and 10th grade pupils in regional school in Hyderabad, India. The objective of the study is to explore whether there is a difference between the richness of lexis in the different proficiency levels. The study is therefore an attempt at proposing lexical variety and density as important indicators of proficiency.

1. 1. Lexical richness and age

Second-language learners vary on a number of individual factors such as personality, motivation, learning style, aptitude and age. It is precisely this last dimension, age, one of the variables that has been most frequently considered in discussions of individual differences in second language acquisition (SLA) (Bialystok, 1997; Bialystok&Miller, 1999; Dulay *et al.*, 1982; Hatch, 1983; Marinova-Todd *et al.*, 2000; Scovel, 2000; Singleton 1989, 1997, 2001; Singleton & Lengyel, 1995). The main concern of age-related research is whether the age at which someone is first exposed to a second language (L2), in the classroom or naturalistically, affects acquisition of that language in any way. As Larsen- Freeman and Long (1991) point out,

some writers claim that SLA is the same process and just as successful whether the learner begins as a child or an adult and /or that adults are really better learners because they start off faster (Ellis, 1985; Flege, 1987). On other hand, it claims that the data obtained in research ambiguous and that adults are at a disadvantage only in a few areas, especially phonology (McLaughlin, 1984). A third group is convinced that younger learners are at an advantage, particularly where ultimate levels of attainment are concerned (Harley, 1986; Patkowski, 1980).

The reasons for this interest in the age issue relate not only to theoretical matters, such as whether children or adults go about acquisition similarly or differently or whether an innate language faculty continues to function beyond a particular maturational point (Martohardjono & Flynn, 1995), but also practical issues such as when L2 instruction should begin at school.

Hence, as many researchers state “Adolescence marks a period of substantial growth traversing many aspects of development, including vocabulary” (Corson, 1997; Berman, 2004). Research in this area often adopts one of two main approaches: investigating quantitative signs of vocabulary growth among adolescents; and enquiry into factors during adolescence which promote vocabulary acquisition. These two approaches have yet to be integrated into the same study. In SLA research, adolescence has been seen primarily as a developmental watershed in which the child’s neurological facility for learning languages is lost or altered (e.g., Scovel, 2000). At the same time, theorists have often portrayed maturing learners’ changing views of self and social context as an important factor in age differences in second language (L2) attainment (e.g., Krashen, 1981). These factors consolidate the pertinence of adolescence period in language acquisition more often than not vocabulary proficiency variation in second language learners.

1. 2. Lexical richness and proficiency in writing (production)

One of the major determinants of the vocabulary used in written production is the vocabulary size of the writer, particularly if the writer is a second language learner with a relatively small vocabulary compared with native speakers. Measures of lexical richness attempt to quantify the degree to which a writer is using a varied and large vocabulary (Laufer & Natin, 1993). A well written composition, among other things, makes effective use of vocabulary. This need not be reflected in a rich vocabulary, but a well-used rich vocabulary is likely to have a positive effect on the reader. Engber (1993) like Linnarund (1986) examined the relationship between various lexical measures of a piece of writing and holistic scores of writing quality. Engber’s study suggests that it is worth helping and encouraging learners to bring their vocabulary knowledge into active use in writing. There are many factors besides vocabulary size that could affect lexical richness in writing. These could include familiarity with the topic, skill in writing, and communicative purpose (Laufer & Nation, 1983). The productive vocabulary is more imperative for second language learners than receptive vocabulary in helping to cope with their communicative use and it is highly expectant to see a relationship between direct measures of learners’ vocabulary size and the richness of vocabulary in their language production.

2. LITERATURE REVIEW

2. 1. Lexical proficiency: Variety and Density

Lexical variation (also known as lexical diversity) is measured by looking at the range of different words used across a text, or in other terms, the extent to which repetition is lacking (Malvern, Richards, Chipere & Durán, 2004; McCarthy & Jarvis, 2013). Several L1 and L2 vocabulary acquisition studies have used lexical variation as a measure, showing overall trends

of significant L1 and L2 lexical development across adolescence (Berman & Verhoeven, 2002; Stromqvist et al., 2002; Malvern et al., 2004; Johansson, 2008; Berman & Nir, 2010; Crossley, Weston, McLain Sullivan & McNamara, 2011).

Spencer project corpora attributed many researches in lexical variation and adolescent development primarily in L1 which also produced many researches in L2 in the same pattern. The Spencer project, initiated with the goal of examining text production abilities of children, adolescents, and adults, recruited participants from the following age groups: 9-10 years, 12-13 years, 15-17 years, and adults (university students ages 20-30 years) (Berman & Verhoeven, 2002). The data were elicited through experimental design, in which participants produced narrative and expository texts on the topic of personal conflict. Seven countries participated (France, Holland, Iceland, Israel, Spain, Sweden, and USA), resulting in seven corpora which allow for reliable cross-linguistic comparisons. Data from Spencer project studies indicate that marked development, in terms of increasing lexical variation levels and lexical density occurred during adolescence. The studies (Stromqvist et al., 2002; Berman & Verhoeven, 2002; Johansson, 2008; Berman & Nir, 2010), covering English, Hebrew, Icelandic and Swedish writing samples, uniformly report on a trend of significant development in lexical variation between 12-17 years, with no significant development occurring in the years immediately before and after (9-12 and 17-adult), except in the Swedish corpus findings where growth in lexical variation levels was observed beyond age 17 (Johansson, 2008). These findings led Stromqvist et al. (2002, p. 53) to conclude that the period of 12-17 years constitute an “important developmental leap” in terms of lexical acquisition. Such findings of lexical variation development may be indicative of other areas of development.

Summarising a body of L1, L2 and language impairment research using the lexical variation measure, Malvern et al. (2004) state that lexical variation is commonly seen as indicative of vocabulary size, and ability in vocabulary use, meaning its development also indicates underlying vocabulary growth. Berman (2007) situates this view within the context of adolescent lexical development, arguing that advanced vocabulary acquisition involves the extension of word knowledge, including acquiring synonyms and understanding polysemous meanings of words, the implication being that individuals have more words at their disposal to convey meanings.

Many studies have thus far examined productive L2 development over time. Vocabulary size in research on productive vocabulary development is commonly measured in terms of how many words learners can produce. It is typically investigated by using free word production tasks (Palmberg, 1987) or de-contextualized productive vocabulary level tests (Nadarajan 2009; Schmitt & Meara, 1998). A pilot study conducted by Palmberg (1987) is perhaps one of the earliest longitudinal studies specifically addressing written vocabulary growth over time. The goal of the study was to examine productive vocabulary development as it took place in an ordinary foreign language classroom setting. Twenty-two Finnish-Swedish bilingual pupils (11 girls and 11 boys, all age 11) learning English as a foreign language participated in the experiment. The pupils were asked to write down as many words as they could possibly think of that began with a given letter within a one-minute period. The data were collected weekly during a 10-week experimental period. This type of a free word production task, in fact, could not provide information about the actual size of an individual’s productive vocabulary. However, the total number of words produced by the pupils during 10 consecutive test sessions, nonetheless, showed that there was a steady increase in word production from session to session.

Schmitt and Meara (1998) showed an increase in participants' vocabulary size by using Nation's Levels test (Nation, 1990), a vocabulary-size test that probes vocabulary knowledge at five frequency levels (words drawn from 2K, 3K, 5K, and 10K frequency lists as well as from the English Academic Word List). The proportion of correctly answered items at each level is taken to be the proportion of total words known at that level, and the totals for each level are added together to reach an estimate of the overall vocabulary size. These authors found that among 95 adult Japanese English L2 learners in foreign language contexts, there was an average 330 word vocabulary gain over the course of one academic year.

In another study Nadarajan (2009) also found an increase in written vocabulary size over time among the participants in his study, regardless of the two different instruction conditions investigated. This author investigated whether explicit vocabulary instruction involving focus-on-form activities would be more beneficial than implicit meaning-based vocabulary instruction in helping learners acquire 20 high-frequency academic words. Laufer and Nation's (1999) Productive Vocabulary Levels Test (PLVT) was used to measure vocabulary gains, and the data were collected at the beginning and the end of the semester from 129 undergraduate English L2 learners in six intact academic writing classes. Findings revealed an increase in the post-test scores for both the implicit group and the explicit instructional group for the 3K word level and words from the Academic Word List (AWL) (Coxhead, 2000). However, contrary to the authors' prediction, the post-test scores for the implicit group showed a greater gain compared to the explicit group, which implies that implicit meaning-based vocabulary instruction might be more effective than explicit vocabulary teaching in helping adult L2 learners enlarge their vocabulary. Nadarajan (2009) pointed out that given the heterogeneous nature of L2 learners, several factors may play important roles in affecting vocabulary development process, such as the level of motivation, the amount of exposure to the target language, in-depth elaboration and familiarity of the words used by instructors, the quality of instruction, and the learners' initial vocabulary level. Nevertheless, all these studies show a predictable growth in proficiency level of L2 learners over a period of time highlighting the value of this study regarding proficiency level across two groups.

Lexical richness in research on productive vocabulary development is generally measured (typically through corpus-based counts) in terms of the number of different types of words learners can produce. However, the specific aspects of vocabulary knowledge that have been targeted are in fact relatively inconsistent across studies. One reason for this is that lexical richness has been operationalized in a variety of ways in the literature. Most commonly used measures of lexical richness include lexical density (i.e. percentage of content words), lexical sophistication (i.e. percentage of less frequent words), and lexical diversity (i.e. type-token ratio).

Laufer (1991), for example, investigated advanced adult L2 English learners' vocabulary development in terms of lexical richness over a two-semester period. The study was carried out in the context where participants received no explicit vocabulary instruction, but were nevertheless exposed to language. Lexical richness was operationalized in terms of four variables, namely lexical variation (LV, defined as the ratio between the different lexemes [individual word families] in the text and the total number of lexemes), lexical density (LD, defined as the percentage of nouns, verbs, adjectives and adverbs in the text), lexical originality (LO, defined as the percentage of lexemes that are used by only one particular writer of the group) and lexical sophistication (LS, defined as the percentage of "advanced words", or words taken from Xue and Nation's (1984) University Word List, a list of 737 base words and over 1,400 derivatives). Free compositions written by the participants were collected at the beginning

and the end of a semester and were later analyzed. The comparison of each of the four variables showed that, with the exception of lexical sophistication, no significant development in productive vocabulary occurred after one academic year for these advanced learners. Examination of individual learners further suggested that learners whose initial vocabulary proficiency levels were below average made progress on each variable in order to function at the average level of their group. However, advanced students who had no problem functioning in their environment and performing their school tasks did not exhibit significant progress. Based on the results, Laufer (1991) proposed the “active vocabulary threshold hypothesis”, suggesting that productive vocabulary development is determined not so much by comprehensible input as by the needs of individual learners for reaching the average level of the group in which they are required to function.

A written corpus of L2 learners was also used in the study conducted by Horst and Collins (2006) to investigate the development of lexical richness in young learners. The participants in this study were beginner-level francophone learners of English (11-12-year-olds) enrolled in an ESL intensive program in the French-speaking province of Quebec. Narrative writing texts produced in response to picture prompts were collected four times at regular intervals of 100 hours of intensive ESL program (approximately 400 hours in total). A computational tool, Vocabprofile (Laufer & Nation, 1995), was used to measure lexical richness by determining the proportions of running words in the corpus which can be found at 1K (1000 most frequent families) and 2K (2000 most frequent families) frequency levels. Surprisingly, these analyses did not reveal the expected increase in use of less frequent words in learners’ written productions after 400 hours of instruction. In other words, after a substantial period of intensive instruction, any changes in the young English L2 beginners’ vocabulary did not seem to entail the acquisition of more advanced words (i.e. less common words).

A spoken corpus was used in a more recent study conducted by Bulte et al. (2008) to investigate oral productive vocabulary development in terms of lexical diversity (i.e. type-token ratio), lexical sophistication (i.e. percentage of less frequent words in the corpus), and lexical productivity (i.e. percentage of content words in the corpus). The target language in this study was French. What was emphasized by the authors was the special “French as a foreign language context” in Brussels, namely, the officially Dutch-French bilingual but predominantly francophone context. The participants, who were Dutch speaking pupils from Dutch secondary schools in Brussels, had had relatively large amounts of naturalistic exposure to French in addition to their four-year regular formal classroom instruction in French. Data from the unplanned oral retelling tasks were collected three times throughout the two-year research period. Notably, 22 different measures were used to analyze the data in order to adequately capture the dynamics of lexical L2 proficiency development over time. Statistical results indicated that learners progressed significantly in terms of lexical diversity, sophistication and productivity after two years of study. However, for the Dutch-speaking pupils, the development of lexical proficiency, particularly in terms of lexical productivity and diversity somehow trailed off after the first year.

Also using a spoken corpus, Crossley et al. (2009) investigated L2 lexical development in the spontaneous speech of six adult L2 English learners at an intermediate level in a one-year longitudinal study in an ESL context. In terms of lexical growth, the authors adopted the Measure of Textual and Lexical Diversity (MLTD, McCarthy, 2006) to measure the type-token ratio of the corpus. According to the authors, MLTD is able to assess differences in lexical diversity even when the texts are considerably different in terms of length. Crossley et al.’s

findings showed that as learners spent time studying the target language, their lexical diversity increased.

To summarize, learner corpora have been commonly used in literature to investigate L2 productive vocabulary development in terms of lexical richness. However, the inconsistencies of research focus in these studies, namely, the specific dimensions of lexical richness investigated, the specific groups of participants studied, and the specific research contexts targeted, in fact, make it difficult to compare the results across studies. Regarding oral productive lexical growth, the findings of Bulte et al. (2008) and Crossley et al. (2009) seem to suggest that in contexts where learners can have adequate exposure to the target language, traceable development with regard to lexical diversity (i.e. type-token ratio) in speaking tasks could be expected from learners over one year of study. In terms of written productive vocabulary development, in contrast, learners showed little progress over time in lexical sophistication (i.e. the acquisition of less frequent words) in the studies by Laufer (1991) and Horst and Collins (2006). It seems that factors including participants' age and proficiency level play an important role in affecting the development of lexical richness. In sum, more research is needed if we are to gain a better understanding of how L2 productive vocabulary develops over time in classroom contexts.

2.2. Measurements: lexical variety

The more varied a vocabulary a text possesses, the higher lexical diversity. For a text to be highly lexically diverse, the speaker or writer has to use many different words, with little repetition of the words already used. The traditional lexical diversity measure is the ratio of different words (types) to the total number of words (tokens), the so-called type-token ratio, or TTR (e.g., Lieven 1978, Bates, Bretherton & Snyder 1988). A problem with the TTR measure is that text samples containing large numbers of tokens give lower values for TTR and vice versa. The reason for this is that the number of word tokens can increase infinitely, and although the same is true for word types, it is often necessary for the writer or speaker to re-use several function words in order to produce one new (lexical) word. This implies that a longer text in general has a lower TTR value than a shorter text, which makes it especially complicated to use TTR in developmental comparisons, e.g., between age-groups, where the number of word tokens often increase with age. Gayraud 2000 compares TTR and the number of word tokens and shows that although the number of word tokens increases substantially with speaker/writer's age, the TTR drops. One consequence of this is that TTR is only possible to use when comparing texts of equal length. In spite of this, TTR is still used for comparing text production, for instance between children's texts, or between various groups with language impairment.

A variant of the TTR measure is the so-called *index of Guiraud*. This measure uses the square root of TTR. Other proposed variants are *Advanced TTR* and *Guiraud Advanced*, for instance used by Daller et al. 2003. The ineffectiveness of the existing lexical diversity measures gave rise to a new measure called D, which has come to be considered an "industry standard" (McCarthy & Jarvis 2010). Malvern and Richards's (1997) D is based on a curve-fitting approach. Its main aim is to find the best fitting curve to model the TTR in the text. Even though based on a sample produced by a single child (Richards, personal communication), this model has gained recognition between researchers (Jarvis 2002; McCarthy 2005). It was, however, soon replaced by a different and more solid, according to the intentions of its authors (Richards, personal communication), procedure which makes use of random sampling and for which a special software was developed, called *vocd*. This procedure is to some extent different from the original approach, which is also the reason why the two are often distinguished in literature with

the former procedure called original D or Da, and the latter called adapted D, Db or vocd-D (Jarvis 2002; McCarthy 2005; McCarthy & Jarvis 2007, 2010).

As we wanted to compare the two groups based on the test result with respect to 8th and 10th grade and their performance in terms of the variable task-type (narrative text, picture description/story writing) an improvised algebraic formula of TTR was used to calculate lexical diversity: $V/\sqrt{2N} = C$. This formula was put forth by Guiraud (1954). V refers to the types of words or ‘types’, N refers to the total number of words or ‘tokens’, C is the value of lexical diversity. This formula can be used to measure lexical diversity in case of varying text length.

2.3 Measurement: lexical density.

The concept of density refers to a kind of complexity that results from the development of words. In other words, this relates to the notion of lexico-grammar in terms of the level of wording in language (Halliday, 1985). Many studies have explored the variation between lexical density and proficiency in L1 and L2 language learners. There are at present several different measurements of lexical density. Originally, it was proposed by Ure (1971) that lexical density should be treated as the proportion of the number of lexical items per the number of running words. This formula was refined by Halliday (1985) as his first approximation to measure lexical density, and was then further developed by O’Loughlin (1995). However, Halliday (1985), the originator of Systemic Functional Linguistics, proposes another formula in determining lexical density of texts based on clauses. He states that “lexical density is the number of lexical items as a ratio of the number of clauses” (Halliday, 1985, 67). In order to measure lexical density, it is necessary to distinguish grammatical items from lexical items and the differences between them. Grammatical items, or ‘function words’, come in a closed system comprising of determiners such as articles, pronouns, most prepositions, conjunctions, some classes of adverb and finite verbs (Cindy & James, 2007; Halliday, 1985b). In contrast, lexical items, or ‘content words’, are named *lexical* as they function in lexical sets, that is to say, they are part of an open system rather than closed sets (Halliday, 1985). Traditionally, nouns, verbs, adjectives and adverbs are the four word classes belonging to lexical items since they have autonomous meaning even in isolation and new members can be added to these categories (T.Le, Yue, & Le, 2011). It is interesting to note that Halliday (1985) uses the term ‘items’ rather than ‘words’ when discussing lexical items and grammatical items, since he argues that they may contain more than one word in the usual sense. Taking several phrasal verbs such as *stand up*, *take over*, or *call off*, each of these consists of two words, a lexical verb and a preposition, but Halliday treats them as a lexical item. This is quite a contrast to the views of Ure (1971), who counts them as two separate words, one being the lexical word *stand*, *take*, or *call*, and the other being the preposition *up*, *over*, or *off* respectively. Halliday (1985) makes clear that in most cases it is unambiguous to define which the lexical item or grammatical item is, however, in some cases it is more difficult to define, English prepositions and certain classes of adverb, *always* and *perhaps* for instance, which fall on this borderline. The variation in the lexical density approves the proficiency level of L1 or L2 learners. This study looks at the proficiency level in English of two different age groups where English is their L2. The study follows Ure’s definition of lexical density, counting the words that have pure lexical properties and separated from the words that possess a more grammatical-syntactic function than the lexical item.

3. RESEARCH QUESTIONS AND HYPOTHESES

1. Is there a difference in the rate of lexical variety between learners of different proficiency level?
2. Is there a difference in the rate of lexical density between learners of different proficiency level?
3. Which kind of lexis features more in both the proficiency level?

4. METHODOLOGY

The participants were asked to write a story in English based on a picture. This method of data collection was pioneered successfully by Jeanine Treffers-Daller and Michael Daller; its main aim is to get samples which would follow the same basic storyline, thus making them more comparable. A picture of an accident scene was selected, as it was considered easy for students to relate to their daily life while the school is located in the centre of a town. Students fulfilled the task as a part of one of their English classes and they were not allowed to use dictionaries or any other materials because the aim of the study was to test the vocabulary they already knew and could use. Two age groups participated in this study, 13 years old (8th grade) and 15 years old (10 grade) students in a regional medium school in Hyderabad, India. All participants use English as their L2 and Telugu as their L1, with no known reading or writing difficulties in both languages. Each group consisted of 5 persons and the text length range was 100 – 300 words.

The present study focuses on free written production. It attempts to measure the breadth of students' vocabulary using different measures which are available. In Nation's (2001) classification the focus is on written form, but not on the meaning or the use of vocabulary. One of the major shortcomings of measures of lexical richness is that they do not take into account how the words are used in the text, whether they are used correctly as far as their meaning in that particular context is concerned, how they function grammatically, whether the text is well formed or even whether it makes sense. These measures only assess the breadth (size) of vocabulary used in the text and they would give the same results for a well formed or for a scrambled text given the same words were used in both. For this reason, judgements about text quality cannot be based solely on measures of lexical richness but other aspects have to be taken into account as well.

5. RESULTS

5.1. Overall performance in lexical diversity and density.

The composition had 2071 tokens and 952 types in total. Two groups were compared. The composition by two groups did not have significant difference in type-token ratio. For the measurement of lexical diversity the TTR (Guiraud, 1954) method was used. The Mean of total 'tokens' of 8th and 10th grade were 179.6 and 234.6, and total Mean of 'types' were 90.2 and 100.2 respectively. The overall type-token ratio was 10.67 for 8th grade and 10.34 for 10th grade which clearly demonstrates that there was no significant difference except a slight difference between two groups in their lexical diversity. Hence, it cannot be claimed that variation exists in the rate of lexical diversity in language proficiency regarding age or class level. It may be stated that choice of words does not necessarily indicate proficiency. What is of importance is whether

the words have any semantic purpose or not. This statement is further complemented in the findings presented below.

Lexical density was calculated in Ure’s (1971) method which describes the percentage of lexical (i.e. content) words in the total number of words in any given text, either written or spoken. The composition had 575 lexical words in total. Comparison between two groups resulted in a significant difference in lexical density. The Mean of 8th grade participates (29.4) was shown higher than 10th grade participants (25.8) showing a Mean difference of 3.6. This result confirms that there is a significant difference between two groups in terms of lexical proficiency in content words. And the age difference and higher lexical density has not any correlation claiming that the 8th grade pupil scored highest lexical density than 10th grade which should be vice-versa, on preconceived idea of the researcher. The findings in this regard proved to be contradictory to the researcher’s initial assumption. It may be due to the nature of the task or the task administration issues. However, regardless of all, the findings do provide some significant light into the fact that there is a difference in content words across different proficiency levels.

5.2 Findings and discussions

Looking at the text length of two groups 10th grade participants used more tokens (Mean 234.6) than 8th grade participants (179.6) which show a significant variation between two groups in their token counts. 10th grade participants used a text length of 200 to 280 words and 8th grade participants’ words were counted in between 140 to 260 words. It also found a variation in ‘types’ count between two groups, 10th grade (mean=100.2) and 8th grade (mean=90.2). It was seen that there is a difference between proficiency levels in producing the text length which serves in substantiating the hypotheses of the researcher that old pupil use more words than the younger pupils. After measuring the individual ‘type token ratio’ it was calculated in group wise to compare the both groups, which produced a mean of 4.76 and 4.61 for 8th and 10th grades respectively. Although, the 10th grade pupil produced more tokens than 8th grade but the use of repeated words was higher in their text which was comparatively less in the 8th grade pupils.

The problem of type token ratio is evident in this calculation. The figure shows there is only slight difference in TTR between two groups, The quantitative finding reveals that the time between years 13 and 15 (grade 8 and 11) do not constitutes a period of significant but slightly different lexical development and language proficiency in the area of lexical variation. Therefore, the research question, Is there a difference in the rate of lexical variety between learners of different proficiency level, was addressed with findings of no correlation between lexical proficiency and class level and surprisingly the younger pupils used more lexical diversity than older pupils.

Table no. 1

Grades	8 th grade	Mean	10 th grade	Mean
				234.6
Tokens	898	179.6	1173	
Types	451	90.2	501	100.2
Type token ratio	23.84	4.76	23.05	4.61

Exploring the second research question, is there a difference in the rate of lexical density between learners of different proficiency level, result shows no correlation in the rate of lexical density and proficiency level of pupils in language learning. 8th grade students had a higher mean (29.4) than 10th grade (M=25.8) which should be resulted visa versa if there is correlation between lexical density and proficiency level. Surprisingly 8th grade pupils exhibit more lexical items in their short texts than 10th grade pupil in their long texts. To measure the lexical density of two groups, researcher used the Ure’s formula, Lexical Density = the number of lexical items * 100/ the total words.

Looking at the third research question, Which kind of lexis features more in both the proficiency level?, a comparison of mean of both groups respect to Noun, Verb, Adjective, and Adverb was applied. As a result 10th grade outperformed 8th grade pupils with respect to the use of noun, adjectives and adverbs except verbs which both group used in a similar quantity.

Table no. 2

Grade 11	Grade 8	Grade 11	Grade 8	Grade 11	Grade 8	Grade 11	Grade 8
Nouns	Nouns	Verbs	Verbs	Adjectives	Adjectives	Adverbs	Adverbs
25	26	17	24	6	2	6	3
30	19	22	17	7	0	5	0
28	35	24	13	4	0	6	4
37	25	18	21	10	1	8	5
26	28	22	26	10	6	5	4
Total words: 146 mean: 29.2	TW:133 M: 26.6	TW: 103 M: 20.6	TW:101 M:20.2	TW: 37 M: 7.4	TW: 9 M:1.8	TW: 24 M:7.4	TW: 15 M: 3.2

From the above table it clearly indicates that 10th grade participants have outperformed the 8th grade participants in the lexical items of adjectives, where the former has a mean of 7.4 and later a mean of 1.8 that shows a difference of 5.6. The use of adverbs in 10th grade is relatively higher than 8th grade in a difference of 4.2. But there did not have significant difference in the use of verbs in both groups (10th:20.6 and 8th: 20.2) and a slight difference of M=2.6 were seen in the use of noun between two groups. Almost participants used noun words like accident, footpath, supermarket, hospital, doctor, etc. The significant difference in mean of adjectives shows the 10th grade participants outperformed the 8th grade. The proficiency level is likely influenced in the using of adjectives in forming the composition more comprehensible and lucid regarding to their exposure to language learning. The more the learner gets exposure in a language the usage of content words differs. Regarding verbs there is a slight difference between two groups; mean of 20.6 and 20.2 was seen in 10th and 8th grade respectively. Evaluating the research question and findings, the 10th grade participants produced more adjectives, adverbs and nouns in their texts than 8th grade pupils. It was seen that two of 8th grade pupil hardly used adjective and adverb in their write ups.

6. CONCLUSION

The study has traversed on the domains of lexis with special focus on density and diversity. It has been an attempt at exploring the most asked question whether quality or quantity holds more worth in language development. The findings of the study suggest that the content words weighs more in importance thereby fulfilling the purpose that proficiency can be measured by the extent of how much a learner knows the usage and function of the words he/she uses. The study does have a lot of limitations with regard to the methodology and the size of the samples. However, it has achieved its purpose in further exploring the function of lexis in the general proficiency level of language learners.

WORKS CITED

- Berman, R. (2004). Between emergence and mastery: The long developmental route of language acquisition. In R. A. Berman (Ed.), *Language development across childhood and adolescence* (pp. 9-34). Amsterdam: John Benjamins.
- Berman, R. A., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language and Literacy*, 5(1), 1-43.
- Coxhead, A. (2011). Exploring specialised vocabulary in secondary schools: What difference might subject, experience, year level, and school decile make? *TESOLANZ Journal*, 19, 37-52.
- Crossley, S., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, 59(2), 307-334.
- Daller, H., Milton, J. & Treffers-Daller, J. (2007b). Editors' instruction: Conventions, terminology and an overview of the book. In H. Daller, J. Milton and J. Treffers-Daller (eds.), *Modelling and assessing vocabulary knowledge*. (pp. 1-32). Cambridge: Cambridge University Press.
- Ellis, N.C. (2001). Memory for language. In P. Robinson (ed.), *Cognition and second language instructions*. Cambridge: Cambridge University Press.
- Halliday, M. A. K. (1989). *Spoken and written language*. Oxford: Oxford University Press.
- Johansson, S. (2007). *Seeing through multilingual corpora: on the use of corpora in contrastive studies*. Philadelphia : John Benjamins
- Laufer, B. (1991). The development of L2 lexis in the expression of the advanced learner. *The Modern Language Journal*, 75(4), 440-448.
- Laufer, B. (1994). The lexical profile of second language writing: does it change over time? *RELC Journal*, 25, 21-33.
- Laufer, B., & Nation, P. (1995). Vocabulary breadth and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307-322
- Linnarud, M. (1986). *Lexis in composition: A performance analysis of Swedish learners' written English*. Lund, Sweden: C. W. K. Gleerup
- Malvern, D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. New York: Palgrave Macmillan.
- McCarthy, M. & Carter, R. (1997). Written and spoken vocabulary. In Schmitt, N., and McCarthy, M (eds.), *Vocabulary: Description, acquisition and pedagogy*. (pp. 20-39). Cambridge: Cambridge University Press.

- Meara, P. (2005). Lexical frequency profiles: A monte carlo analysis. *Applied Linguistics*, 26(1), 32-47.
- Nadarajan, S. (2009). The effect of instruction and context on L2 learners' Vocabulary development. *Electronic Journal of Foreign Language Teaching*, 6(2), 177-189.
- Palmberg, R. (1987). Patterns of vocabulary development in foreign-language learners. *Studies in Second Language Acquisition*, 9(2), 201-219.
- Schmitt, N., & McCarthy, M. (Eds.) (1997). *Vocabulary: Description, acquisition, and pedagogy*. Cambridge: Cambridge University Press.
- Stromqvist, S., Johansson, V., Kriz, S., Ragnarsdottir, H., Aisenman, R., & Ravid, D. (2002). Toward a cross-linguistic comparison of lexical quantain speech and writing. *Written Language & Literacy*, 5(1), 45-67.
- Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren & J. L. M. Trim (eds.) *Applications of linguistics: Selected papers of the second international congress of applied linguistics, Cambridge 1969*. (pp. 443-452). Cambridge: Cambridge University Press.