Empirical Studies on Correlations between Lexical Knowledge and English Proficiency of Chinese EFL Learners in Mainland China over the Past Two Decades

Yao Zhou      Zhongxin Dai
College of Foreign Languages, North China Electric Power University
No 2, Beinong Road, Huilongguan, Changping District, Beijing 102206, China

Abstract
Knowledge of English vocabulary contributes to the learner’s proficiency of English as a foreign language, but how the learner’s lexical knowledge behaves in the contribution. Researchers in mainland China have conducted studies of various kinds in order to find out how the learner’s lexical knowledge correlates with his proficiency. This article reviews the empirical studies on correlations between lexical knowledge and English proficiency over the past two decades. The correlations concerned in this review refer mainly to the relationships between lexical knowledge and the overall English proficiency, between the breadth and depth of lexical knowledge and the skills of listening, reading and writing in English as a foreign language.

Keywords: correlation, empirical study, lexical knowledge, English proficiency, Chinese EFL learner

1. Introduction
Vocabulary is the fundamental basis for a language learner to understand and employ a language. Just as Hatch (1983:74) stated, “...when our first goal is communication...it is the lexicon that is crucial to make basic communication possible.” The vocabulary size of language learners directly affects the development of their listening, speaking, reading and writing skills. However, grammar teaching and learning has been occupying the central position in traditional English teaching in mainland China. Vocabulary teaching and learning is of secondary importance. In 1980, the British psycholinguist Meara published an article named “Vocabulary Acquisition: A Neglected Aspect of Language Learning”. Since then, lexical research has been attracting public attention and has become the fastest growing area in second language acquisition research in terms of research output and publications. Vocabulary is no longer a victim of discrimination in second language acquisition research. After decades of neglect, vocabulary is now generally recognized to be central to the second language acquisition process. Vermeer (1992: 147) pointed out, “Knowing words is the key to understanding and being understood. The bulk of learning a new language consists of learning new words. Grammatical knowledge does not make for great proficiency in a language.” Gass & Selinker (1994: 270) also believed that “the lexicon may be the most important component for learners”. Stæhr (2008) observed, “Vocabulary knowledge is generally assumed to be a good predictor of language proficiency in a second or foreign language, and it has long been recognized that vocabulary size in particular plays a crucial role for L2 learners’ communicative competence in English.” It is obvious that vocabulary plays an important role in language learning process, which, to some extent, would influence learners’ language proficiency.

Over the last two decades, a handful of studies have been conducted on the issue of the correlation of lexical knowledge and English proficiency in mainland China, and tremendous findings have been achieved. This paper focuses on the empirical investigations, conducted in mainland China, on the correlations between lexical knowledge and English proficiency of learners of English as a foreign language. It aims at projecting an overall picture of the empirical studies on the correlations.

2. Methodology and Data Collection
We first searched the papers from China National Knowledge Infrastructure (CNKI), which is one of the largest databases of academic papers published in mainland China, by key words of “vocabulary” and “English proficiency”, “vocabulary” and “reading”, “vocabulary” and “writing”, “vocabulary” and “listening”, and “vocabulary” and “speaking”. Then we selected papers published over the past two decades (from 1996-2016) in important journals, most of which are from the so-called core journals listed in the Catalog of Core Journals in China, published by Peking University. Through careful examination, we finally selected about 25 papers, which can be considered as empirical studies on correlations between lexical knowledge and English proficiency, and the skills of listening, speaking, reading and writing. The table below shows the main journals reviewed:
Researchers often investigate language proficiency from those two aspects. For instance, Lv (2004) investigated the vocabulary size and its influence on English achievement as well as its relationship to the depth of lexical knowledge. This study involved 1610 freshmen in a national key university in southwestern China, and 200 subjects were randomly chosen according to their scores on a placement test. The results reveal that vocabulary size does not have a great influence on English achievement, only 34.7 percent of which can be predicted by vocabulary size. The correlation between vocabulary size and English proficiency is just 0.336, but vocabulary size cannot predict listening achievement. What’s more, the influence of vocabulary size on English achievement varies between learners with different levels of English achievement; and the learners’ depth of lexical knowledge becomes greater with their increase in vocabulary size. But for learners with different vocabulary sizes, there exists disproportionate development in their depth of lexical knowledge.

Li (2007) got a different result from his research. The result of his research indicates that the vocabulary size has a correlation of 0.086 with listening proficiency, 0.283 with reading proficiency, and 0.319 with writing proficiency, and in general, vocabulary size has a high correlation (0.39) to the overall language proficiency. To investigate the correlations between breadth and depth of vocabulary knowledge and second language proficiency levels, Li employed 168 second year non-English majors as research participants. Findings show that both breadth and depth of vocabulary knowledge are effective predictors of second language proficiency levels; that the depth of vocabulary knowledge makes a unique contribution to the prediction of second language proficiency, over and above the prediction afforded by vocabulary size, especially in cloze and writing; and that depth and breadth dimensions of vocabulary knowledge are highly, and positively, correlated.

Yang (2008) investigated the relationships among the breadth and depth of vocabulary knowledge and English proficiency levels of 81 Chinese polytechnic non-English majors. According to Pearson correlation analysis (test), there is linear relationship among participants’ vocabulary size, depth of vocabulary knowledge, and the English scores, which is 0.618, 0.631 and 0.765 respectively (P=0.000<0.1). Yang also found that individual differences in vocabulary depth serve as the most effective predictors for English proficiency of the average and high achievers, while reading is the predominant predictor for the underachievers’ English proficiency. This finding proves Lv’s (2004) conclusion that the vocabulary size affects English achievement and the development of depth of vocabulary knowledge varies for different levels of learners.

Yang and Yang (2012) made a research on the depth of the English knowledge of freshmen and sophomores in their university. They found that there is a significant positive correlation between the depth of vocabulary knowledge and the comprehensive English proficiency. That is to say, the better command of the depth of vocabulary knowledge, the better comprehensive English proficiency of the learners.

Zhao and Song (2015) surveyed 5,030 non-English major freshmen and explored the relationship between their vocabulary knowledge and language competence. The research findings indicate that the relationship between their depth of English vocabulary knowledge and language competence (r=0.609, p<0.01) is higher than that between their vocabulary breadth knowledge and language competence (r=0.478, p<0.01). This finding is consistent with Li’s (2007) result. Besides, the depth of vocabulary knowledge can better predict language competence than the breadth of vocabulary knowledge, which has a difference of 14.2%; in view of language skills, both vocabulary breadth and depth knowledge have better prediction about students’ reading and writing.

All of the above studies on correlations between lexical knowledge and English proficiency indicate that there are positive relationships between vocabulary size or the depth of lexical knowledge and English proficiency. The research subjects are all college students. Their language proficiency can be predicted, to some extent, by examination of their breadth and depth of vocabulary knowledge. As for their specific language skills, different studies vary in different aspects.
4. Lexical Knowledge and Reading

It is commonly believed that lexical knowledge is one of the most important factors influencing reading comprehension in second language research. A considerable number of studies have found significant correlations between vocabulary knowledge and reading comprehension competency for EFL learners from different proficiency levels. For example, Qian’s (1999, 2002) studies show that there are high and positive correlations among the vocabulary size, depth of vocabulary knowledge and reading comprehension. The depth of vocabulary knowledge can also contribute significantly to the prediction of L2 reading comprehension. Qian’s research is not comprehensive since it does not cover all respects of vocabulary depth.

Yang and Deng (1996) took four months to investigate the influence of vocabulary teaching on the reading of students of science and engineering. They found that in the control classes the students’ vocabulary and reading scores improve significantly (0.01<p<0.05), the increase of the vocabulary and reading scores of the students in the experimental class is very significant (p<0.01). Therefore, it is obvious that vocabulary knowledge and reading scores are closely related. Li (2003) also did an experiment on the correlations between vocabulary knowledge and reading comprehension, and his conclusion reveals that the correlation between the breadth of vocabulary knowledge and reading comprehension is highest (0.62). That is to say, EFL learners with large vocabulary size usually perform well in reading comprehension. He also found that some aspects of the depth of the vocabulary (e.g. word context, syntax) are worth paying attention to while some other aspects of the depth of the vocabulary (e.g. polysemy) are not since the latter does not seem to have much significant relationship with reading comprehension.

Zhang and Qiu (2006) also investigated the relationship between breadth and depth of vocabulary knowledge and reading comprehension. They drew a conclusion that breadth and depth of vocabulary knowledge are positively correlated with reading comprehension, which proves the results of Yang and Deng’s (1996) study. The conclusion also shows that depth of vocabulary knowledge is more closely correlated with reading comprehension than breadth of vocabulary. This result differs from Li’s (2003) study, which believes that vocabulary size is more closely correlated with reading comprehension. This difference may result from different experimental participants. The former one (Zhang and Qiu) studied English majors and the latter (Li) focused on non-English majors.

Gong (2006) investigated 60 higher vocational students. He divided them into two groups, one group (A) that passed CET 4 and another (B) that didn’t. Linear regression analysis shows that the depth of vocabulary knowledge of Group A has obvious predictive power of reading comprehension, which means that the richer the vocabulary knowledge is, the stronger the reading ability of the students is. This result confirmed Yang et al’s (1996) conclusion: the level of English vocabulary and reading ability has positive correlation. Li (2015) investigated the relationships among metacognitive knowledge, vocabulary size and EFL academic reading, based on 548 non-English major sophomore students in China. The findings reveal that Chinese tertiary EFL readers have a good command of 2000-word level and approach 3000-word level. Vocabulary size significantly influences EFL reading comprehension ability, which explains 19% of the total variation of reading (p<001). The result once again confirms Yang et al’s (1996) and Gong’s (2006) findings.

All of the above studies on correlation between vocabulary knowledge and reading ability suggest that there are strong positive correlations between vocabulary size and reading comprehension. And there also exist close and positive inter-correlations among vocabulary size, depth of vocabulary knowledge and reading comprehension. Moreover, compared with the vocabulary size, the depth of vocabulary knowledge correlates more closely with reading comprehension. This result can be found in all the studies except Li’s (2003).

5. Lexical Knowledge and Listening

Vocabulary knowledge and listening also has a strong correlation. Kelly (1991) argued that vocabulary knowledge is the main obstacle to successful listening comprehension for EFL learners. Chinese researchers have carried out in recent years many investigations on the relationship between vocabulary knowledge and listening comprehension. Zhang (2011) conducted an empirical research on sophomore English majors to examine the relationship between lexical knowledge and listening comprehension of TEM-4. He found that the breadth and depth of lexical knowledge significantly correlates with listening comprehension. The breadth of lexical knowledge accounts for 27% of the variance of listening comprehension, 24% of the variance of dictation and different variance of other parts; while the depth of lexical knowledge explains 2% of the variance of listening comprehension and dictation respectively. Therefore, it is obvious that the correlation coefficient between breadth of lexical knowledge and listening comprehension is larger than that of the depth of lexical knowledge and listening comprehension. Later, in 2014, Zhang carried out a further investigation on the roles of short-term memory, working memory and lexical knowledge in L2 listening and reading comprehension. The results proved the result of his previous study that vocabulary knowledge has medium significant correlation with listening comprehension. The correlation coefficients among productive vocabulary knowledge, vocabulary size and the listening skill are 0.52 and 0.47 (P<0.01) respectively. The vocabulary size accounts for 31.4% of
the variance of listening comprehension, while the productive vocabulary knowledge explains 3.1%. By hierarchical regression analyses, vocabulary knowledge can account for 39.6% of the variance of listening skill. Those findings reveal that listening skills improve with the increase of the breadth and the depth of vocabulary knowledge.

Zhang’s (2011, 2014) studies mainly focused on English majors. Wang et al. (2011) conducted an empirical research with 95 non-English major sophomores to explore the relationship between the breadth and depth of vocabulary and grammatical knowledge and listening comprehension. Their findings also prove that the lexical knowledge (breadth and depth) and listening comprehension have significant positive correlations, and that depth of vocabulary knowledge of all the samples and that of the samples with small vocabulary size all explain the variance in listening comprehension (45.3% and 57.1% respectively). The results accord with Zhang’s (2011, 2014) findings.

The aforementioned studies reveal that correlations between vocabulary knowledge and listening comprehension are significant, and the lexical knowledge can explain the variance of listening skills. These results are different from Lv’s (2004) investigation, whose findings indicate that vocabulary knowledge cannot predict the level of listening skills and there is no significant correlation between the two factors. The difference in the methods and the procedures may be the main reason for the different results.

An experiment carried out by Du and Qiu (2015) explored, by word recognitions, the learners’ attentional focus in listening comprehension. The analysis of the words recognized indicates that the scores of the students who focus on content words, nouns and prepositions in prepositional phrases recognition are higher than the scores of those who focus on function words, verbs and prepositions in verb phrases. In addition, the research also indicates that learners with a more balanced noun / verb recognition profile tend to have high listening comprehension scores. It once again proves Zhang’s (2011, 2014) and Wang et al.’s (2011) conclusions that vocabulary knowledge and listening comprehension are closely correlated.

6. Lexical Knowledge and Speaking

Relatively fewer investigations are conducted on the relationship between vocabulary knowledge and the skill of speaking in mainland China. The reason for it might be that speaking skill is hard to measure and the experiment procedure is difficult to control. We only found one paper in our review. This paper investigates the influence of lexical knowledge on L2 oral production of sophomores majoring in English. The correlation analysis shows that the correlation coefficient of vocabulary size, productive vocabulary knowledge and oral English proficiency are 0.44 and 0.39, at or near average correlation (0.4≤r≤0.6). The findings also indicate that both the vocabulary size and the productive lexical knowledge can predicate the variance in L2 oral production, and that the role of vocabulary size is more significant (Zhang, 2015).

7. Lexical Knowledge and Writing

Lexical knowledge is a very important factor in EFL writing.Lots of studies have been carried out on the relationship between learners’ use of vocabulary and the quality of their writing. Ma and Wen (1999) investigated the relationships of L2 learners’ linguistic variables to L2 writing ability. They found that L1 writing ability, L2 speaking ability and L2 productive vocabulary together can explain 73% of the variance of L2 writing ability. Moreover, the influence of L2 productive vocabulary (0.43) is larger than L1 writing ability and L2 speaking ability, which suggests that the proficiency of English productive vocabulary plays a vital role in English writing ability. If learners have a larger vocabulary size, they can use more and better expressive words to manifest the theme of the composition, hence increased writing quality.

Liu (2003) analyzed 57 second-year college students’ writing samples. Nation’s (1990) 10,000 word level test was used to measure the productive vocabulary size of the students, and their vocabulary used in timed composition was analyzed. This study adopted Nation’s (1995) Lexical Frequency Profile, and used VocabProfile software to analyze the subjects’ lexical richness in their writing. The results indicate that vocabulary size has no immediate effect on the writing score, but it indirectly affects the writing score by influencing the text length. Because large vocabulary size often exerts an effect on text length, it influences writing quality. Besides, learners with a large vocabulary tend to use relatively more sophisticated words and less K1 words (the first 1,000 most frequent words). Liu (2004) also made an attempt to investigate whether productive vocabulary size affects writing quality with writing strategy training as a moderator variable, and whether strategy training can improve writing quality when L2 learners do not have a large vocabulary. Sixty sophomores from two randomly chosen classes at the same proficiency level participated in this study. The control group with a larger vocabulary size did not receive strategy training. In contrast, the experimental group with a smaller vocabulary size was trained to use writing strategies such as planning. The results show that productive vocabulary size has no significant effect on writing quality, and that high writing quality can be attributed to the combination of a large vocabulary size and writing strategy use. The results also reveal that writing strategy can do much to compensate for the subjects’ smaller vocabulary size. However, when the
vocabulary size reaches a certain threshold, the writing quality may increase with larger vocabulary size.

The lexical usage and lexical problems would also affect the quality of writing. Liu et al. (2009) analyzed the composition corpus of 67 engineering students, and found that the students’ notional words are limited and that delexicalized verbs are less used in writing. Hence, they concluded that vocabulary teaching should be enhanced to improve students’ writing skill. Furthermore, some scholars focused on students’ lexical problems and their solutions in writing compositions. Xu and Ding (2010) used think-aloud and stimulated recalls to trace and describe six English majors’ lexical-problem-solving strategies in timed writing process. They discovered that students’ writing proficiency differs with different grades in retrieving a lexical candidate: third-year students would further employ strategies to fine-tune expressions to achieve lexical preciseness, stylistic appropriateness and avoid repetitive use of a certain word in the context, while first-year students would immediately employ strategies to repair lexical forms. Thus it is obvious that the ability to use lexical strategies would eventually influence writing proficiency. Yu (2011) carried out a research through a questionnaire survey and an interview to investigate the difficulties which 88 English majors experienced in extracting the productive words while writing compositions. He found four lexical problems in writing: lack of topic-related words and cohesive expressions, wrong use of common core words, monotonous diction and slow extraction of words from the mental lexicon. Those major problems would affect writing qualities. Bai and Dai’s (2013) research aimed to understand the contribution of different dimensions and frequency bands of lexical knowledge to the ability of reading and writing by testing the vocabulary knowledge, reading and writing ability of 136 first-year English majors. The vocabulary tests assessed three dimensions of word knowledge: word recognition, spelling and collocation. The first two dimensions cover 2000, 4000 and 6000 words of three word frequency bands while the last covering only 2000 words frequency band. Correlation and multiple regression analyses of the data show that collocation knowledge of 2000 words frequency band made a unique contribution above all other variables to the quality of writing, explaining 24.8% of the variance. The study also suggests that strengthening students’ collocation knowledge of high frequency words can achieve a better command of English language.

It is interesting to see that the word length in the English writing of middle school students in China is different from that in the United States. Li (2013) made a contrast research and found that the average word length of Chinese middle school students’ writings is bigger than that of American students. It is because Chinese students like to use longer low-frequency words than American students. They tend to use fewer function words which mainly consist of two letters. This study reflects the problem in the usage of middle school students’ vocabulary in our country, and the importance of vocabulary teaching in writing.

As regards the development of lexical richness or lexical diversity in English writings, Wan (2010), Wang and Zhou (2012), Zhu and Wang (2013) carried out empirical studies from different perspectives. Wan (2010) investigated the development of lexical diversity in English majors’ writings from three perspectives: lexical variation, lexical sophistication and lexical errors. He found that as learners further their English language study, they improve both their lexical variation and lexical sophistication, while their error types and error tokens are on the increase. Of all the major errors, the error in using the articles is the most frequent one, and improvement in verb uses can be expected. Wang and Zhou’s (2012) longitudinal study concerned the developmental features of lexical richness in English writings. They investigated 30 non-English majors, focusing on such aspects as lexical variation, lexical sophistication, lexical density, and lexical errors. The results show that as the participants’ English level rises, a steady increase can be expected in lexical variation, lexical sophistication and lexical density. This finding proves the finding of Wan (2010). Students make fewer lexical errors as they make progress, but spelling errors remain the most serious. This result is different from Wan’s (2010). It may be caused by the different participants. Moreover, Wang and Zhou (2012) once again proved that the relationship between lexical knowledge (including lexical variation, lexical sophistication, lexical density, and lexical errors) and the writing quality is positively correlated. Zhu and Wang (2013) explored the developmental features of lexical richness in English writing, based on a self-built corpus of 120 English compositions on a same topic written by 30 Chinese English majors throughout their four-year study period. Their study indicates that there exist a steady yet not straight-line progress and a plateau phenomenon in the students’ lexical variation during their four-year English learning. The study concludes that as the students improve English proficiency, their acquisition patterns and developmental paths of the multi-dimensions of lexical richness in English writing are convergent but with salient differences.

8. Conclusion
This review of the empirical studies on correlations between lexical knowledge and English proficiency of Chinese EFL learners enables us to shed some light on the research area of lexical knowledge and English proficiency in mainland China. With knowledge of corresponding research abroad and in-depth analysis of the reports, we draw the following conclusions.
(1) Most of the researchers draw conclusions from their research that vocabulary size affects English achievement to some extent but the influence differs with learners with different levels of English achievement.
Some of the researchers concentrate on the depth of vocabulary knowledge, and find that the depth of vocabulary knowledge and the comprehensive English proficiency are positively correlated. The depth of vocabulary knowledge can better predict language competence than vocabulary size. Five reports in the review on the correlations between lexical knowledge and English proficiency mainly focus on non-English majors. Participants need to be diversified. Further research is needed to study the correlations between lexical knowledge and overall English proficiency of English majors, high school students, etc.

(2) Most Chinese researchers show great interest in the influence of vocabulary knowledge on reading and writing ability, but fewer researchers investigate the influence on listening and speaking competency. The general consensus is that reading comprehension is strongly affected by vocabulary size. Correlations between reading competency and the breadth and the depth of lexical knowledge exist, but the degree of correlation varies with different participants and researchers (see Li, 2003; Zhang and Qiu, 2006). Further explanatory research is desired to account for the causes of the contradictions in the reports. Writing quality is strongly affected by lexical knowledge. Vocabulary knowledge in English writing can be observed from such dimensions as lexical variation, lexical sophistication, and lexical density (Wang and Zhou, 2012). Findings show that these measurable dimensions of lexical knowledge are correlated with writing qualities.

(3) Only a few researchers conducted research on the correlations between vocabulary knowledge and listening comprehension and speaking competency as it might be hard to design and control the experiment. Only Zhang (2015) carried out an in-depth study on the influence of lexical knowledge on L2 oral production. Four reports investigated listening skills. The studies indicate that vocabulary knowledge can predict, to some extent, the level of listening skills, except Lv’s (2004) investigation, which found no significant correlation between the two factors. This contradiction is worthy of further exploration. More empirical studies are expected to test the relationship between vocabulary knowledge and speaking competency. Vocabulary can be divided into receptive and productive vocabulary, with the former contributing to the skills of reading and listening and the latter to speaking and writing. Lexical knowledge also involves the changing process from lexical knowledge to lexical skills. More in-depth research will be expected to explore those aspects.

References


