A Comparison of Productive Vocabulary in Chinese and American Advanced English Learners' Academic Writings^{*}

Xixiang Lou

Faculty of Foreign Languages, Nanjing Normal University, Nanjiang, China; Department of Foreign Languages and Literatures, Zhangzhou Normal University, Zhangzhou, China Email: Lou_xx@163.com

Guanghui Ma

Faculty of Foreign Languages, Nanjing Normal University, Nanjiang, China Email: mgh@njnu.edu.cn

Abstract—A comparison has been made of productive vocabulary in some normal university English majors' theses in China and American final-year undergraduates' papers. The research demonstrates that with family as the measurement unit, Chinese students proportionally use fewer words of the 1st, 8th to 10th 1000 frequency level words than American students, while in terms of the 2nd to 4th 1000 frequency band words, Chinese students use more than American students and there is no difference in the usage of the words of the rest word lists. In terms of lexical words complexity: American students repeatedly use the 1st 1000 high-frequency words in a higher proportion, use more inflectional forms than Chinese students but use fewer derivational forms than Chinese students: Use more inflectional and derivational forms than American students; use more inflectional and derivational forms than American students, but use low-frequency words repeatedly in a greater proportion, more inflectional forms in the same proportion as American students.

Index Terms-productive vocabulary, richness, lexical word complexity

I. INTRODUCTION

Second or Foreign Language learning process is often described as a developmental continuum of interlanguage from the state of being utterly ignorant to that of complication without necessity of arriving at the perfect stage. Such a incremental development trend should also be embodied in vocabulary research, because target language native speakers are most obviously different from foreign language learners in their use of greater number of vocabulary, especially in free speech or writing. (Laufer, 1994)

The distinguishment between Productive Vocabulary and Receptive Vocabulary was made only by intuition, but now has been widely accepted and attempted to be explained theoretically (e.g. Henrisen. 1999; Waring 2002). Comapred with Receptive (Passive) vocabulary research, Productive (Active) Vocabulary research came up later. Some research has been made on SL learners' Productive Vocabulary at home and abroad (Laufer, 1994 1995, 1998; Laufer & Nation, 1995, 1999; Leń ko-Szymań ska, 2009; Webb, 2008, 2009; Bao Gui , 2008; Li Yao & Shen Jinhua, 2009; Li ZhiXue & Li Jingquan, 2005; Lu Min, 2008; Tan Xiaochen, 2006, 2007; Wan Lifang, 2010), and great achievements have been made, which has deepened people's understanding of Production Vocabulary. However, rarely any consensus has been reached in such an aspect and there are still some gaps in such a field, for the variability of productive vocabulary ability and the influence of various factors in research.

On the one hand, few studies have been made on normal university English majors, one of the most important FLA groups, still less any attention has been paid to the productive vocabulary usage in English major thesis, as their virgin academic writing. Normal university English undergraduates are to set foot on teaching, and in the future process of teaching, they will have to be engaged both in teaching and scientific research. In the process of teaching, they will make use of their own English knowledge and skills to cultivate and improve their students' English ability. Accordingly, these future teachers' highest English level will affect that of the students they teach. This consists in the significance of the teachers' English level. As preparation for the future scientific research, the normal university English majors' thesis writing is one of the most important indicators of their ultimate and highest writing ability, the research on which will no doubt contribute to the evaluation of their level of language usage and academy and will suppy FL teaching with powerful benchmarks. On the other hand, there are some inconsistent conclusions about the

^{*} Research grant from the Social Science and Humanities Council of the Ministry of Education of China in 2012, entitled "Study of the Relational Dimensions of L2 Speaking Ability" (12YJC740033)

differences between Chinese EFL learners and English native speakers in terms of Productive Vocabulary development research among domestic research.(e.g. Gui Lin, 2010; Li Zhixue & Li Jingquan, 2005; Wen Qiufang, Ding Yanren & WangWenYu, 2003).

In this paper, a comparison will be made betwee English major graduates of a Chinese normal university and American college students in their vocabulary quality of productive academic writing texts.

II. THEORETICAL FRAMEWORKS

A word is the minimal unit which can be used independently to form sentences or discourses. In the counting of words, different units can be adopted to different requirement. A token refers to the running word or occurrence of words. The number of tokens refers to the total number of word forms, which means that each occurrence of the same word is counted as one individual word. A type refers to different word forms in a text, and the number of types is the total number of the different word forms, sot a word which is repeated many times is counted only once (Read 2000 Assessing Vocabulary p. 18). In vocabulary studies, the base and inflected forms of a word are collectively known as a lemma (Read 2000 p. 18). So a collection of words such as *work, works, working, worked*, comprising a root form and the most frequent regular inflections, is known as a lemma (Daller et al 2007 p. 3). A lemma consists of a headword and some of its inflected and reduced (n't) format. Usually, all the items included under a lemma are of the same part of speech. The English inflections consist of plural, third person singular, present tense, past tense, past participle, -ing, comparative, superlative and possessive (Francis and Kucera, 1982: 461 cited from Nation 2001 P. 7) A word family consists of a headword, its inflected forms, and its closely related derived forms (Nation 2001 P. 9). A word family comprises the base word plus its inflections and its most common derivatives (c.f. Thornbury 2002 p.3). So a word family headed by *employ* includes *employs*, *employed*, *employed*, *employee*, *employee*, *employees*, *employeer*, *employeers*, *employeers*, *employeers*, *employees*, *employees*, *employees*, *employees*, *employeers*, *employees*, *employees*, *employeers*, *employeers*, *employees*, *employees*

Lexical richness is defined as the quality of vocabulary usage of a text and will vary with the depth and breadth of vocabulary knowledge (Nation and Webb, 2010). The measurement of lexical richness is to quantify the degree of the variety and amount of words of writers(Laufer & Nation, 1995). Nation and Webb(2010) think that the comprehensive measurement of vocabulary richness should supply some or all of the following contents:

1. provide the number of different tokens, types, lemmas, and families in a text

2. show the word frequency of all items in the text

3. list the number of encounters with each type, lemma, and family in the text

4. display the mean number of words per sentence, paragraph, and text

5. show the number of errors and categorize the errors accordingly(in other words, incorrect spelling, derivation, meaning, use)

6. show the extent to which different affixes are used

7. indicate the extent to which semantically-related words are used together

8. provide the proportion of the major parts of speech (nouns, verbs, adjectives, and adverbs)

9. indicate how well words are used together (to what extent do words appear in frequent collocations or formulaic sequences?)

10. display the time it takes to use a word, phrase, or sentence in speech or writing

11. display the mean number of words read per minute (p: 248-249)

Token-Type Ratio (TTR) refers to the total number of tokens of words in a text divided by that of types used here as the indicator of the average of repetition of words in a text. It is the reciprocal of Type-token ratio, a much used measure of lexical diversity that is the number of different words in a sample of speech or writing divided by the total number of words (David et al., 2004 p. 192).

Type-Lemma Ratio (TLR) refers to the total number of types of words in a text divided by that of lemmas, used here as the indicator of the average number of inflected word in a text.

Lemma-Family Ratio (LFR) refers to the total number of lemmas of words in a text divided by that of families, used here as the indicator of the average number of derivative word in a text.

III. RESEARCH DESIGN

A. Subjects and Materials

107 subjects involved in this research are the final-year English majors from a provincial normal university. Materials for research are their theses written about linguistics and language teaching, with the cover, the contents part, acknowledgement, references part and appendix part, Chinese annotation and some long direct citations cut off. After purification, the size of the corpus is 501,055 tokens, with an average of 4,440.86 tokens, the shortest 3,148 and the longest 6,961.

The materials for comparison are 93 essays from the Michigan Corpus of Upper-level Student Papers (MICUSP), written about English language or Linguistics (19 linguistics and 74 English language). The writers of these articles are all final-year college students (Final Year Undergraduate). The size of the corpua is 196,816 tokens.

B. Instrument

The software *Range BNC* (Heatley, Nation, & Coxhead, 2002.) has been used in the study to compare a text against 14 vocabulary lists to see what words in the text are and are not in the lists, and to see what percentage of the items in the text are covered by the lists. These 14 wordlists have been made by Paul Nation from the British National Corpus (BNC) based on the breadth, frequency and dispersion of its words. Each of the 14 wordlists contains about 1000 words (lemma) in BNC, and words from the first list to the 14th list occur less and less frequent in BNC.

RANGE BNC provides a table as follows which shows how much coverage of a text each of the base lists provides.

WORD LIST	TOKENS/%	TYPES/%	FAMILIES
One	3544/81.36	616/61.72	428
two	412/9.46	180/18.04	133
three	82/1.88	55/ 5.51	52
four	58/ 1.33	39/ 3.91	34
five	24/0.55	18/ 1.80	17
six	129/ 2.96	10/ 1.00	10
seven	11/0.25	6/ 0.60	6
eight	16/0.37	8/ 0.80	7
nine	8/0.18	8/ 0.80	7
ten	5/0.11	5/ 0.50	5
11	3/ 0.07	3/ 0.30	3
12	4/ 0.09	4/0.40	3
13	4/ 0.09	3/ 0.30	3
14	4/ 0.09	3/ 0.30	3
not in the lists	52/1.19	40/4.01	?????
Total	4356	998	711

It is shown in the above table result that 3,544 of the running words in the text are of base list one and they make up 81.36% of the total running words in the text. Moreover, these 3,544 running forms are of 616 types accounting for 61.72 of the total number of word types in the text and are of 428 families. In the word list column, *one, two, three*, etc. refer to each of the base lists and a statistics list of word not in all those base lists----*not in the lists*. With the support of such software, the statistics of words in each text can be gained, and the average of the statistics for words of each base list for Chinese and American students' texts can also be computed. Then a comparison of word usage (indicating productive vocabulary) of Chinese and American students can be made.

C. Treatment of Materials

In the study, *Range BNC* is used to analyse the texts to show the average proportion of different frequency level words used in the texts to the total number of words of the texts with *family* as the measurement unit to compare the differences in the proportion of the corresponding level words between Chinese and American students' written materials. In general, the greater is the writers' vocabulary, the more he tends to use low frequency words.

Additionally, diversity of words of different frequency level words has been compared between Chinese and American students with the following indicators: a) token-type ration (TTR); b) type-lemma ratio(TLR); c) lemma-family ratio (FLR).

IV. RESULTS AND DISCUSSION

A. The Word Proportion Comparison between Chinese and American Students' Writing

For a comparison of proportion of words used in writings from the 14 base lists by Chinese and American students, the averages of proportion of words from the 14 base lists used in Chinese and American students' writings are listed in table 1:

TABLE 1	
THE PROPORTION OF WORD FAMILIES FROM THE 14 BASE LISTS USED IN CHINESE AND AMERICAN THESES (NCHINESE=107, NAMEI	RICAN=93)

DI	Chinese	American			DI	Chinese	Chinese		American	
DL	per	s.d.	per	s.d.	- BL	per	s.d.	per	s.d.	
1	54.06	4.40	58.31	6.23	8	1.27	.41	1.47	.63	
2	19.67	1.41	16.39	1.96	9	.79	.32	1.05	.45	
3	7.68	1.19	6.46	1.46	10	.57	.33	.70	.54	
4	5.71	1.03	5.10	1.07	11	.61	.28	.70	.52	
5	3.04	.81	2.97	.99	12	.41	.25	.47	.39	
6	1.98	.62	2.18	.89	13	.34	.21	.42	.41	
7	1.56	.54	1.52	.73	14	.25	.19	.25	.26	

Notes: BL=base list; per=percentage

It is shown in table 1 that Chinese students use fewer word families from base list 1, 6, 8, 9, 10, 11, 12 & 13 in proportion, use more from base list 2, 3, 4, 5, & 7 and use the same proportion of word families from base list 14. Such

a tendency can also be seen in the following bar chart.



The independent t-test results show that it is only on base list 1, 2, 3, 4, 8, 9 and 10 that the word family percentage in the theses of Chinese and American students differs from each other significantly, i.e. Chinese students use more word families on base list 2,3 and 4 than American students (t=-13.432, -6.514, -4.077; p=.000, .000, .000; df=164.72, 198, 198), while on base list 1, 8, 9 and 10, Chinese students use fewer word families than American students (t=5.497, 2.620, 4.536, 1.981; p=.000, .010, .000, .049; df=162.515, 155.323, 162.763, 149.593). Such a result is different from that of Li Zhixue and Li Jingquan (2005) which shows that Chinese students use less words proportionally from the first base list than American Students (the same as this research), while on the rest base lists, Chinese students are not different from American ones significantly.

In order to compare with the result of Li Zhixue and Li Jingquan (2005), the results of the base list 5 to 14 have been blended into a low frequency word list. The independent t-test result shows that American students have used a greater proportion of low frequency words than Chinese students (t=1.99, p=.049, df=153.62). Such a result is quite different from that of Wen Qiufang (2003), i.e. Chinese students use a 14% greater proportion of the first base list words than Native students, a 4% greater proportion of the second base list words and a less than 5% greater proportion of the rest two base list words.

The difference maybe results from four aspects:

First, foreign language learners are inherently different from native speakers, because they are mainly depending on the classroom teaching materials for language input. The frequency of word occurrence in their textbooks can hardly be the reflection of the natural language context of native speakers (Daller, Milton, & Treffers-Daller, 2007). Usually, the frequency of foreign language learners' encountering new words is far lower than native speakers. It is no doubt that foreign language learners are usually different from native speakers in word usage.

Secondly, there is a greater demand on college or university students' theses writing in Chinese teaching syllabus. Usually a thesis writing course has been set as a selective or compulsory one, and most of times, some teachers will be assigned to instruct some specific students' thesis writing. They will keep on coaching, supervising and urging their students' thesis writing and revising all the way. Naturally, the theses will have higher qualities than expected, word usage included.

Thirdly, the theses used in this research were written in the semester when the university was accepting the educational assessment from the Ministry of Education. Therefore, teachers had a far stricter demand on students' thesis writing. Most of the theses have been read through by teachers and revised by students for many times. At the same time, students have tried to cite directly or indirectly from many sources. Naturally, students' thesis writing will seem to have a higher quality. Such an idea is supported by Li Zhixue and Li Jingquan's research (2005).

Fourthly, there are some differences in corpus data used in these researches. Li Zhixue and Li Jingquan's research (2005) used daily free writings, which are quite different from those used for this research in academy. It is natural for students to use a greater proportion of high-frequency words similar to American students do in their writings. Wen Qiufang (2003) has used timed writings in exams as data for analysis. For a high mark in an exam, usually, students will try to avoid the usage of difficult words---usually low-frequency words---to avoid making mistakes. So it is no wonder that the students have overused high-frequency words in her research. So using timed writings to assess students' vocabulary may underestimate their ability and proficiency. In this research, the open, untimed, freely-written theses have been used for analysis for they are more academic and suitable for formal words. Therefore, they have a greater proportion of relatively low-frequency words.

B. The Comparison between Chinese and American Students in Word Variation

In the research, a frequency-decreasing wordlist of Chinese and American students' corpora has been respectively made. From the top 200 words of these two wordlists and based on frequency order, 70 shared words have chosen, and these words which have not or are not inflected or derived forms are mainly function words (articles, prepositions, conjunctions, pronouns) and very few content word (numbers). In the process of corpora neating, these 70 shared words have been eliminated from two corpora to gain a more accurate statistical result.

It takes two steps to get the number of lemma or lexeme: First of all, lemmatize the language materials of texts in corpora, i.e. replace inflected words with their base form. Then, analyze the corpora with Range BNC for the second time, and the number of word types gained is the number of lemma. In the study, the data for the base list from 3 to 14 have been blended into one low-frequency base list data. The statistics of the comparison are shown in table 2.

STATISTICS OF TTR, TLR AND LFR FOR CHINESE AND AMERICAN STUDENTS WRITINGS								
Statistics	Chinese(N=107) American(N=93)							
Indicators	Min.	Max.	М	s.d.	Min.	Max.	М	s.d.
TTR for base list 1	2.35	4.63	3.10	.46	3.13	74.30	14.61	12.00
TTR for base list 2	1.48	3.81	2.47	.48	1.20	9.00	1.76	.86
TTR for low	1 32	4 39	2.12	59	1.12	9.00	1 77	93
frequency base list	1.52	4.57	2.12	.57	1.12	2.00	1.77	.,,,
TLR for base list 1	1.16	1.33	1.25	.03	.09	4.00	.32	.54
TLR for base list 2	1.04	1.30	1.14	.04	1.00	1.27	1.11	.05
TLR for low	1.01	1 16	1.07	03	07	1.12	1.05	03
frequency base list	1.01	1.10	1.07	.05	.91	1.12	1.05	.05
LFR for base list 1	1.11	1.25	1.18	.03	1.00	1.30	1.12	.04
LFR for base lis2	1.02	1.23	1.11	.04	.97	1.25	1.08	.05
LFR for low	1.02	1 1 5	1.08	02	1.00	1.15	1.07	03
frequency base list	1.02	1.15	1.00	.02	1.50	1.15	1.07	.05

TABLE 2	
OF TTD TID AND I ED FOR CURRENT AND AN	(FRIGAN) OF

It is indicated in table 2 that on average the TTR on the base list 1 of American students (14.61) is far greater than Chinese students (3.10), i.e. American students repeatedly use each of the first base list words 14.61 times, while Chinese students use 3.10 times. The independent sample t-text indicates that such a difference is statistically significant (t=-.9243, p=.000, df=92.24). The standard deviation (s.d.) of Chinese students' result (.46) is far smaller than that of American students' (12.00) and this indicates that there is a greater difference among individual American students than Chinese students. This can also be seen in the statistics of maximum and minimum. For American students, the minimum is 3.13 and the maximum is 74.30. There is a great difference in-between. For Chinese students, the minimum is 2.35, the maximum is 4.63, and the difference is not so great. On the second base list and the low frequency base list, Chinese students' TTR's (2.47 and 2.12) are greater than that of American students' (1.76 and 1.77). The independent sample t-test indicates that such differences are statistically significant (for the base list 2: t=7.04, p=.000, df= 138.81; for low frequency base list: t=3.11, p=.002, df= 151.70), which suggests that Chinese students repeatedly and significantly use more words on base list 2 and low frequency base list than American students.

It is also shown in table 2 that on average the TLR on the base list 1 of American students (4.00) is far greater than Chinese students (1.18), i.e. American students use 4 inflected forms for each of the first base list words on average, while Chinese students use 3.10. Such a difference has been proved to be statistically significant by the independent sample t-test (t=-16.35, p=.000, df=92.61). The standard deviation (s.d.) of Chinese students' result is (.03) is similar to that of American students (.04), which indicates that there is no great difference among both American and Chinese individual students. On the base list 2 and low frequency base list, Chinese students' TLR's (1.14 and 1.07) are greater than that of American students' (1.11 and 1.05). The independent sample t-test indicates that such differences are statistically significant (for the base list 2: t=5.04, p=.000, df= 198; for low frequency base list: t=6.01, p=.000, df= 198), which suggests that Chinese students use significantly more inflected words on base list 2 and low frequency base list than American students.

It is also demonstrated in table 2 that on average the LFR on the base list 1 of Chinese students (1.18) is far greater than American students (1.12), i.e. Chinese students use 1.18 derived forms for each of the first base list words, while American students use 1.12. Such a difference has been proved to be statistically significant by the independent sample t-test (t=-10.97, p=.000, df=151.82). The standard deviation (s.d.) of Chinese students' result (.46) is far smaller than that of American students (12.00) which indicates that there is a greater difference among American individual students than among Chinese individual students. This can be seen in their minimums and maximums. For American students, the minimum is 2.35, and maximum is 4.63. The difference is not so great. On the base list 2 and low frequency base list, Chinese students' LFR's (1.11 and 1.08) are greater than that of American students' (1.08 and 1.07). The independent sample t-test indicates that difference on the base list 2 is statistically significant, while on the low frequency base list, the difference is not statistically significant (for the base list 2: t=4.54, p=.000, df=153.98; for low frequency base list: t=.92, p=.361, df=162.75), which suggests that Chinese students use significantly more derived words on base list 2 than American student and on the low frequency base list, they are similar.

V. CONCLUSION

A research has been made on the proportion and diversity of different base list words used in American and Chinese Academic writings. It is found that as far as the breadth of word usage is concerned, there are statistically significant differences between Chinese students and American students in the percentage of words on base lists 1, 2, 3, 4, 8, 9 and 10. That is to say, Chinese students use more word families on base list 2, 3 and 4than American students, while use fewer than American students on base list 1, 8, 9 and 10.

On the whole, Chinese students use fewer words on both high-frequency words and low-frequency words than American students. As for word diversity, American students more repeatedly use words on base list 1 than Chinese students, and use more inflected forms, but Chinese students use more derived forms. Chinese students more repeatedly use words on base list 2 than American students and use more inflected forms and derived forms. Chinese students more repeatedly use words on low frequency base list than American students and use more inflected forms and the usage of derived forms is similar to American students.

There are some shortcomings in the controlling of the consististency of topics of theses and the length of texts, so the conclusion drawn in this research should be to further test.

REFERENCES

- Daller, H., Milton, J., & Treffers-Daller, J. (2007). Editor's introduction. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), Modelling and Assessing Vocabulary Knowledge. Cambridge: Cambridge University Press.
- [2] Heatley, A., Nation, I.S.P. and Coxhead, A. (2002). RANGE and FREQUENCY programs. http://www.vuw.ac.nz/lals/staff/Paul_Nation (accessed 10/2/2011).
- [3] Henriksen, B. (1999). Three Dimensions of Vocabulary Development. SSLA, 21, 03–317.
- [4] Laufer, B. (1994). The Lexical Profile fo Second Language Writing: Does it Change over Time?. *RELC Journal*, 25(2), 21-33
- [5] Laufer, B. (1995). Beyond 2000: a Measure of Productive Lexicon in a Second Language. In L. Eubank, L. Selinker & M. Sharwood-Smith (Eds.), *The Current State of Interlanguage: Studies in Honor of William E. Rutherford* (pp. 265-272). Amsterdam John Benjamins Pub Co.
- [6] Laufer, B. (1998). The Development of Passive and Active Vocabulary in a Second Language: Same or Different?. *Applied Linguistic*, 19(2), 255-2271
- [7] Laufer, B., & Nation, P. (1995). Vocabulary Size and Use: Lexical Richness in L2 Written Production. *Applied Linguistics*, 16(3), 307-321.
- [8] Laufer, B., & Nation, P. (1999). A Vocabulary-size Test of Controlled Productive Ability. Language Testing, 16 (1), 33-51.
- [9] Leńko-Szymańska, A. (2009). How to Trace the Growth in Learners' Active Vocabulary? A Corpus-based Study. In B. Kettemann & G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis*. Beijing: World Publishing Corporation.
- [10] Nation, I. S. P. (2001). Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.
- [11] Nation, I. S. P., & Webb, S. (2010). Researching and Analyzing Vocabulary. Boston: Sherrise Roehr.
- [12] Webb, S. (2008). Receptive and Productive Vocabulary Sizes of L2 Learners. SSLA 30 79-95.
- [13] Webb, S. (2009). The Effects of Receptive and Productive Learning of Word Pairs on Vocabulary Knowledge. *Regional Language Centre Journal* 40(3), 360-376.
- [14] Bao G. (2008). A Survey of the Development of Lexical Richness in L2 Compositions from a Multidimensional Perspective. Computer-assisted Foreign Language Education in China (123), 38-44.
- [15] Gui L. (2010). A Contrastive Study of Lexical Proficiency between L1 and L2 Compositions via Computerized Assessment. Foreign Language Teaching and Research (bimonthly), 42 (6), 445-450.
- [16] Li Y., & Shen J.H. (2009). A Study of Productive Vocabulary of First-year Non-English Majors. *Foreign Language Research* (3), 117-120.
- [17] Li Z.X., & Li J.Q. (2005). An Empirical Study on the Productive Vocabulary of Advanced Chinese EFL Learners—Based on A Contrastive Analysis of Model Compositions of Advanced Chinese EFL Learners and American University Students. *Shandong Foreign Languages Teaching Journal* (5), 56-59.
- [18] Lu M. (2008). The Developmental Characteristics of the Productive Vocabulary Breadth: A Research based on English Majors' Written Materials. Foreign Language Learning Theory and Practice (2), 10-15.
- [19] Read, J. (2000). Assessing Vocabulary. Cambridge: Cambridge University Press.
- [20] Tan X.C.. (2006). A study of Chinese English learners' productive vocabulary development. *Foreign Language Teaching and Research* (bimonthly), 38(3), 202-207.
- [21] Tan X.C. (2007). A Study of English Learners Development of In-depth Knowledge on Productive Vocabulary. *Foreign Language Education*, 28(2), 52-56.
- [22] Thornbury, S. (2002). How to Teach Vocabulary. Harlow: Pearson ESL.
- [23] Wan L.F. (2010). An Empirical Investigation into Lexical diversity of Chinese English Majors TEM Writings. Foreign Language World (1), 40-46.
- [24] WARING, R. (2002). Scales of Vocabulary Knowledge in Second Language Vocabulary Assessment Retrieved 5th Oct., 2011, from http://www.robwaring.org/papers/various/scales.htm
- [25] Wen Q.F., Ding Y.R., & Wang W.Y. (2003). Features of Oral Style in English Compositions of Advanced Chinese EFL learners: An Exploratory Study by Contrastive Learner Corpus Analysis. *Foreign Language Teaching and Research* (bimonthly), 35(4), 268-274.

Xixiang Lou was born in Henan, China in 1969. He received his M.A degree in Foreign Language Literatures from Henan Nomal University, China in 2004. He is currently a lecturer in the Department of Foreign Languages and Literatures, Zhangzhou Normal University, Zhangzhou, China. His research interests include English Language Testing, SLA, and Corpus Linguistics. At present he is studying for his Ph.D in Linguistics and Applied Linguistics in Nanjing Normal University.

Guanghui Ma was born in 1957. He graduated from Nanjing University in 1998 and was conferred with the Doctor Degree in Literatures and now is a professor of the Faculty of Foreign Languages, Nanjing Normal University. His researches are focused on SLA, Second Language Writing and Corpus Linguistics.