

# Analysis of Prefabricated Chunks Used by Second Language Learners of Different Levels

Liwei Zhu

Tianjin Polytechnic University, Tianjin, China

**Abstract**—This paper first briefly introduces the previous studies on second language learners' chunk use. Then it gives the definition and classification of prefabricated chunks. In part two, it explains the process of the research design by introducing three important criteria—frequency, accuracy and variation to measure student's ability in chunk use. In part three, the results of the research are discussed in detail which shows that English learners' ability to use chunks is positively correlated with their language level. Second year in college is the most important year for second language learners to acquire prefabricated chunks.

**Index Terms**—prefabricated chunks, writing, frequency, accuracy, variation

## I. INTRODUCTION

One's ability to use prefabricated chunks is an important criterion to measure one's language level. Recent years, research in the chunk use by second language learners has been on the increase. Various papers and studies have been published in this field which includes literal reviews and empirical studies. There are some common features concerning these chunk researches: (1) The research content involves chunk using, chunk teaching, chunk defining and measuring of one's chunk ability. (2) The research method is mainly corpus-based and empirical, for example, the relevance between chunk identifying and learners' English level (Huang Qiang 2002, Diao Linlin 2004), analysis of chunk types in second language learners' writing (Xu Xianwen 2010) etc. These studies have arrived at similar conclusions: (1) There is a close relationship between second language learners chunk level and their listening speaking reading and writing abilities. (2) Most of them learners couldn't use English chunks very well, whether they are English majors, non-English majors, lower grade students or higher grade students, especially in writing and speaking. These studies provide the theoretical base for the future research in terms of methodologies. However, there is one issue which has been neglected: There is a common intuition among both English teachers and second language acquisition researchers that second language learners of different levels tend to use these chunks differently. Nevertheless, there has been very few studies as to how different they are or which sections these differences lie in. Therefore, this paper aims to provide some detailed analysis of chunk use by second language learners of different levels in college.

## II. DEFINITION AND CLASSIFICATION OF PREFABRICATED CHUNKS

There have always been great controversies about the definition and classification of prefabricated chunks. However, researchers through years of studies and observations reached a consensus: prefabricated chunk is a language structure that combines the features of both vocabulary and grammar; it performs a specific language function. Simply speaking, prefabricated chunk is a set term which may include one or more words. According to the data retrieved from corpus, prefabricated chunks are the meaningful collocations in the text that reach a certain frequency. Nattinger (1992) categorized the lexical phrases into 4 kinds: (1) poly word (so to speak, by the way,) (2) institutionalized expressions (how are you, have a nice day) (3) phrasal constraint (as far as..., a...ago) (4) sentence builder (my point is that..., not only...but also...). Biber (1999), according to the academic terms he studied in research papers, classified the chunks into 12 kinds: (1) noun phrase +phrase fragment (2) noun phrase +attribute post modifier (3) prepositional phrase +of phrase fragment (4) other prepositional phrase fragment (5) it +verb phrase/adjective phrase fragment (6) passive verb+ prepositional phrase fragment (7) be+ noun phrase/adjective phrase fragment (8) verb phrase +that clause (9) verb/adjective +phrase fragment (10) adverbial clause fragment (11) pronoun/noun +be(+...) (12) other expressions

This paper is to take both categorizations into consideration, leaving out the complex parts which are difficult to retrieve from corpus, and recategorize the studied chunks.

## III. RESEARCH DESIGN

### A. Research Questions

The study aims to answer the 3 questions as follows (1) Among different age groups, what are the differences of chunk use in second language learners of different levels (2) Among the same age groups, what are the differences between students with different language levels? (3) Are there any patterns or features of these usages?

### B. Research Subject

The studied second language learners are divided into 2 groups: One group is the second year non-English major undergraduate in college; they are all from school of computer and software engineering. These 60 students include 5 females and 55 males. Based on their English test scores of Chinese College Entrance Examination, their CET4 (College English Test band 4) and their final term examinations, this group is subdivided into 2 groups. One group contains 15 students with higher English scores; the other group contains 15 students with lower English scores. The second group students are the 60 first year graduate students from grads school. They are different majors. According to their scores in graduate entrance exam, they are also subdivided into 2 groups: 15 students with higher English scores and 15 students with lower English scores.

### C. Corpus Source and Data Collection

The corpus of this study comes from the written homework assigned by the English teachers. The homework is an argumentation entitled *Living On Campus or Off Campus*. Students are supposed to write at least 150 words in English. The composition turned in through Email in the form of txt file.

### D. The Indexes to Measure the Chunk Use

This study measures the second language learners' chunk use with three indexes: fluency, accuracy and variation.

#### 1. Frequency

This index comes from the study by Wolfe-Quintero in 1998. Frequency means the token number of the chunks used in a speech. In this case, it refers to the token number of prefabricated chunks in a written text. Each different chunk is considered to be a type; a type has different tokens. For example, *to look up*, *looked up*, *looking up* are regarded as one type with three different tokens. Since token can better explain second language user's ability to use chunks, it is taken as one index. More over, due to the different word number used in the composition, the total token number should be standardized. The formula is: the total token number  $\div$  total word number  $\times$  1000, i.e. the total chunk number in every 1000 words. (Here the total number refers to the total number in each composition)

#### 2. Accuracy

Accuracy refers to the grammaticality in chunk use; the formula is total number of the correct chunk  $\div$  total chunk number  $\times$  1000. what is needed to point out here is that the identified errors include intralingua errors and interlingua errors. The former kind of error means that the structure of the chunk is incomplete or incorrect; the latter one means that the chunk itself is right but inappropriately used with its context. The grammatical errors which involve the number of noun, verb tense and agreement are not included in this study. For example:

- (1) If the house catch fire, what can he do? (grammatical error)
- (2) All of us *were addicted* in the story. (intralingua error)
- (3) He doesn't care about *that he has no money or power* (interlingua error)

The chunk "catch fire" in the first sentence belongs to the grammatical error, thus not included in the study.

#### 3. Variation

Variation refers to the ratio of token number and word number. Here, the token number means the token number of chunks. It is one of the criteria to decide whether the students are good at chunk using. Due to the different length of the material, the study adopts the formula proposed by Wolfe-Quintero in 1998: type number  $\times$  type number  $\div$  word number. Here the type number refers to the correct type number of the chunks used in every text. The bigger the variation is, the larger the number is, of the chunks students have mastered. Contrarily, students' using the same chunks repeatedly means they are limited in chunk volume.

### E. Corpus Analysis

There are three steps to analyze the corpus: (1) Calculate one by one the fluency, accuracy and variation of the chunks in each written text. (2) Compare the differences between the two groups of students in the above three indexes by using the repeated measures. (3) Compare the differences between students with high English scores and students with low English scores within the same group and analyze the reasons of these differences.

### F. Chunk Defining and Retrieving

The defining of chunks is based on the Longman Modern English Dictionary (2003) combined with English native speaker's intuition. The standards are as follows:

- (1) combination of two or more than two words
- (2) If the above combination appears in the dictionary, it is considered to be a chunk.
- (3) If there is an ambiguous term, it is left to the foreign teacher to decide.

Last, the data are calculated with Excel.

## IV. RESULT AND DISCUSSION

### A. The General Differences of Chunk Use

There are some differences between undergraduate group and graduate group in fluency, accuracy and variation. (see Table 1)

TABLE I.  
THE GENERAL CHUNK USE OF THE TWO GROUPS

Group	Fluency		Accuracy		Variation	
	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
Undergraduate	86.84	20.89	.78	.14	16.07	6.27
Graduate	108.18	22.70	.90	.09	21.30	7.52

TABLE II.  
PAIRWISE COMPARISON OF THE TWO GROUPS

	Fluency	Accuracy	Variation
Mean difference	21.34	.12	5.23

The result of the pairwise comparison shows that the graduate group is better than the undergraduate group in all three sections: fluency, accuracy and variation. The graduate group is composed of a number of best students from college. They must be equipped with strong second language ability to be admitted to grad school. According to the interview of these students, 58 of them have passed CET 6 (College English Test Band 6.). By contrast, the undergraduate group has only finished one year learning of English in college. Only 28 students passed CET 4; none of them passed CET6. However, according to table 1, within each group, there is a large standard deviation existing in fluency, accuracy and variation. This means that even in the same group, students tend to develop differently in their ability to use chunk. The reason might be that there are at least one or two students in each group who are much better or much worse than the other students, which is normal considering the individual factors.

This shows that one's ability to use chunks is positively correlated to one's second language ability.

#### B. *The Differences between High Score Students and Low Score Students*

In order to know the internal differences within the same group, the researcher divides the group into 2 subgroups: 15 students with high English scores and 15 students with low English scores. The following table is the analysis of the two groups.

TABLE III.  
THE CHUNK USE OF THE UNDERGRADUATE AND GRADUATE GROUP

	Fluency				Accuracy				Variation			
	undergraduate		graduate		undergraduate		graduate		undergraduate		graduate	
	M	S	M	S	M	S	M	S	M	S	M	S
H	114.24	11.32	141	14.86	.92	.03	.99	.01	23.93	5.89	31.47	8.8
L	61.24	10.92	86.76	6.21	.60	.12	.80	.09	9.37	2.25	15.75	1.54

M: mean, S: Standard deviation, H: high score group, L: low score group

From table 3, we see that the students with high English scores use chunks better than students with the low English scores; no matter it is in the undergraduate or graduate group. This result shows that student's chunk level is determined by his or her second language level. The higher the score, the stronger their language ability to use chunk is. That means student's second language level is positively correlated with his or her chunk level.

There are three reasons to explain this. First, the conceptions about chunks are different. Learners with higher English scores have a clear conception about prefabricated chunks; they think it extremely important to memorize these set terms. Contrarily, learners with lower English scores couldn't give a clear explanation of it, let alone its functions. In most cases, they tend to solve their learning problems by resorting to grammatical rules. Second, in terms of acquisition, group with high English scores acquire the chunks mainly through recitation and imitation of the phrases they learned in class or saw in the books or movies. However, the students with low English scores merely pay attention to the vocabularies, ignoring their collocations with the context words.

Another interesting fact that comes to the researcher's attention is that when we are observing the data of learners with higher scores, the standard variation of the graduate group is bigger than that of the undergraduate group in all three sections except accuracy. When we compare the learners with low English scores in graduate group and the learners with high English scores in undergraduate group, we may find that the mean value in the undergraduate group is bigger than that of the graduate group in all three sections (fluency, accuracy and variation). On the other hand, the standard variation of high score students in undergraduate group is bigger than low score students in graduate group. There might not be solid reasons to explain this. However, we may infer the in undergraduate group, there might be several students who are extremely good, even better than the graduate students; or we may boldly guess that the best students in the undergraduate group are better than lower score students in graduate group in terms of their language ability. This is probably because second year in college is the year when student make greatest effort to learn English. Thus, their English level is high. According to some previous studies (Qi Yan, 2010; Zhang Jianqin 2004), second year is the critical year for Chinese learners to acquire English. In this year, they make the biggest progress because students have to make preparations for CET 4 and TEM 4. After that, some students are relieved of the pressure and stop working hard. Thus, they couldn't make big progress. This might explain data from the table.

## V. CONCLUSION

The conclusions of this study are as follows: first, English learners' ability to use chunks is positively correlated with their language level. The higher their English level is, the stronger their ability to use language. Second, second year in college is the most important year for second language learners to acquire prefabricated chunks. Third, these differences in chunk use are due to various learning strategies adopted by students. High level students have a clear understanding of what chunk is. They will acquire English chunks through memorizing English chunks in class and watching English movies or reading English novels after class. The intentional and unintentional acquisition work together. Low level students attach importance only to vocabulary and grammar, thus neglecting English chunks.

What should be noted here is that this study has limitations. For example, the sample is relatively small; the result needs further testifying and improving. The future study will widen the scope of the sample and detail the definition of prefabricated chunks so as to further analyze the pattern of chunk use and enhance second language learner's understanding of their inter-language development.

#### APPENDIX I. DATA OF UNDERGRADUATE STUDENTS' CHUNK USE

Note: C refers to Chunk; W refers to word;

No.	WType	CToken	CType	CorrectC	Frequency	Accuracy	Variation
1	200	18	17	12	90.00	.67	16.06
2	207	22	21	16	106.28	.73	20.05
3	151	12	12	9	79.47	.75	12
4	392	48	38	28	102.04	.70	36.1
5	315	29	28	26	92.06	.90	27.03
6	171	15	15	13	87.71	.87	15
7	160	20	19	15	125.00	.75	18.05
8	396	37	37	30	93.43	.81	37
9	168	19	19	15	113.00	.79	19
10	251	20	19	18	79.68	.90	18.05
11	121	8	8	7	66.11	.88	8
12	151	13	13	12	86.09	.92	13
13	328	21	20	16	64.02	.76	19.05
14	175	16	15	12	91.42	.75	14.06
15	258	18	18	16	69.76	.89	18
16	205	16	15	12	78.04	.75	14.06
17	191	16	14	12	83.76	.75	12.25
18	131	4	4	3	30.53	.75	4
19	161	11	11	9	68.32	.82	11
20	205	18	16	13	87.80	.72	14.22
21	190	17	16	14	89.47	.82	15.06
22	144	12	12	12	89.33	1.0	12
23	207	12	12	10	57.97	.83	12
24	223	21	20	18	89.68	.90	22.05
25	215	17	15	14	79.06	.82	15.05
26	154	12	12	8	77.92	.67	12
27	165	14	13	12	84.84	.86	12.07
28	244	24	23	19	98.36	.79	22.04
29	172	9	9	8	52.32	.89	9
30	165	19	19	18	115.15	.95	19
31	194	18	18	13	92.78	.72	18
32	182	16	16	13	87.91	.81	16
33	273	35	32	31	128.2	.89	29.26
34	160	14	13	9	87.5	.64	12.07
35	244	17	17	10	69.67	.59	17
36	119	9	9	7	75.63	.78	9
37	156	13	12	9	83.33	.69	11.08
38	193	11	11	7	56.99	.64	11
39	145	9	8	4	62.04	.44	7.11
40	180	16	16	8	88.88	.50	16
41	143	20	20	19	139.00	.95	20
42	295	34	28	28	115.25	.82	23.06
43	188	21	20	18	111.70	.86	19.05
44	180	20	19	15	111.11	.75	18.05
45	215	27	25	24	125.58	.89	23.15

46	92	6	6	3	65.21	.50	6
47	162	11	11	10	67.90	.91	11
48	172	15	15	4	87.2	.27	15
49	212	16	16	13	75.47	.81	16
50	258	21	21	20	81.39	.95	21
51	141	17	17	15	120.56	.88	17
52	240	11	10	7	45.83	.64	9.09
53	121	9	9	5	74.38	.56	9
54	184	18	18	16	97.82	.89	18
55	262	23	22	18	87.78	.78	21.04
56	222	15	14	12	67.56	.80	13.07
57	241	21	20	15	87.13	.71	11.25
58	194	20	19	18	103.09	.90	18.05
59	171	17	16	13	99.41	.76	15.06
60	219	20	19	19	91.32	.95	18.05

## APPENDIX II. DATA OF GRADUATE STUDENTS' CHUNK USE

No.	WType	CToken	CType	CorrectC	Frequency	Accuracy	Variation
1	242	23	19	18	95.04	.83	15.70
2	198	19	19	16	95.96	.84	19
3	149	15	15	15	100.67	.10	15
4	226	21	21	19	92.92	.90	21
5	224	23	23	18	102.68	.57	23
6	223	26	26	24	116.59	.92	26
7	217	17	17	12	78.34	.71	17
8	175	16	16	11	91.43	.69	16
9	182	20	18	17	109.89	.85	16.2
10	406	50	50	47	123.00	.94	50
11	229	26	24	22	113.54	.92	22.15
12	230	25	24	24	108.70	.96	23.04
13	259	33	32	31	127.41	.94	31.03
14	156	17	17	16	108.97	.94	17
15	208	27	27	27	129.81	.10	27
16	217	27	26	26	124.42	.96	25.04
17	218	23	21	23	105.50	.10	19.17
18	286	30	25	26	104.90	.87	20.83
19	198	19	18	18	95.96	.95	17.05
20	148	18	18	18	121.61	1.0	18
21	186	24	24	23	129.03	.96	24
22	186	24	24	23	129.03	.96	24
23	201	22	22	19	109.45	.86	22
24	308	35	34	33	113.64	.94	33.03
25	244	20	20	19	163.93	.95	20
26	244	20	29	19	81.97	.95	20
27	283	23	23	22	81.27	.96	23
28	283	26	25	24	91.87	.92	24.04
29	182	22	22	18	120.88	.82	22
30	220	19	18	17	86.36	.89	17.05
31	158	16	15	14	101.27	.88	14.06
32	181	22	20	22	121.55	1.0	18.18
33	153	20	20	18	130.72	.90	20
34	190	19	18	14	100.00	.74	17.05
35	184	18	18	17	97.83	.94	18
36	145	22	21	20	151.72	.91	20.05
37	162	13	12	10	80.25	.77	11.08
38	165	18	17	17	109.10	.94	16.06
39	321	50	50	49	155.76	.98	50
40	262	43	41	43	164.12	1.0	39.09
41	172	28	28	27	162.80	.96	28
42	222	34	29	32	153.15	.94	24.74

43	213	27	26	23	126.76	.85	25.04
44	180	21	21	20	116.67	.95	21
45	159	17	17	17	106.92	1.0	17
46	188	27	26	27	143.62	1.0	25.04
47	246	19	17	16	77.24	.84	15.21
48	190	23	22	23	121.05	1.0	21.04
49	425	40	40	39	94.12	.98	40
50	224	22	22	22	102.68	1.0	22
51	223	20	19	19	89.69	.95	18.05
52	166	15	15	13	90.36	.87	15
53	222	25	24	24	112.61	.96	23.04
54	169	22	22	21	130.18	.96	22
55	196	21	21	21	107.14	1.0	21
56	170	17	17	15	100.00	.89	17
57	216	17	17	16	78.70	.94	17
58	170	17	17	15	100.00	.89	17
59	187	21	20	20	112.30	.96	19.05
60	196	18	18	16	91.84	.89	18

#### ACKNOWLEDGEMENT

This work is supported by a grant from the Humanities and Social Sciences Projects in Tianjin universities and colleges in 2012 organized by Tianjin Education Commission. (Project name: *Study of the Chunk Use by Second Language Learners of Different Levels*; project number: 20122207).

I would like to give my sincere thanks to all those who have helped me with the thesis. Without their constant support and guidance, this thesis would never have been completed. First, I should express my heartfelt gratitude to my husband, Shang Chuang, who has offered his huge help in processing the complicated and time consuming data. I should also thank my colleague Professor Li Li, who has offered me valuable suggestions in the academic studies. In preparation of this paper, she has spent much time reading through each draft and provides me with inspiring advice. Without her insightful criticism, the completion of this thesis wouldn't have been possible.

I also want to show my thanks to my colleagues who have done me a great favor in collecting data for my paper.

Finally, I'd like to express my love and gratitude to my parents who have been helping me out of difficulties and supporting me without a word of complaint. This work is dedicated to them.

#### REFERENCES

- [1] Biber D, Johansson S, Leech G, Conrad S & Finegan E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.
- [2] Butler C S. (2003). Multi-word sequences and their relevance for recent models of functional grammar. *Functions of Language*, (10), 179-208.
- [3] Diao Linlin. (2004). Research on the chunk use by Chinese undergraduates. *Journal of The People's Liberation Army Institute of Foreign Languages* (4), 35-38.
- [4] Ding Yanren. & Qi Yan. (2005). Relevance study of chunk use and students' spoken and written English ability. *Journal of the People's Liberation Army Institute of Foreign Languages* (3), 49-53.
- [5] Han Baocheng. (2010). *Statistics in foreign language teaching and research*. Beijing: Foreign Language Teaching and Research Press.
- [6] Huang Qiang. (2002). Empirical study of the vocabulary collocation acquisition of the advanced English learners. *Journal of The People's Liberation Army Institute of Foreign Languages* (4), 76-79.
- [7] Lewis M. (1993). *The Lexical Approach: The State of ELT and a way forward*. Hove, UK: Language Teaching Publications.
- [8] Lewis M. (1997) *Implementing the lexical approach: Putting theory into practice*. London: Language Teaching Publications.
- [9] Li Wenzhong & Zhang Jun (2004). Verb-Noun collocation pattern and error analysis based on COLEC. *Foreign Language Teaching*, (4), 30-32.
- [10] Li Wenzhong. (2004). Interlanguage collocation and analysis of second language learners' strategy based on COLEC. *Journal of Henan Normal University*, (5), 202-205.
- [11] Ma Guanghui. (2009) Study of the chunk use in the limited time writing by English majors. *Foreign Language Teaching and Research*, (1), 54-60.
- [12] Nattinger James R. & DeCarrico Jeanette S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- [13] Pu Jianzhong. (2003). The Colligation, collocation and chunk in English vocabulary teaching. *Foreign Language Teaching and Research*, (6), 438-445.
- [14] Qi Yan (2010). The longitudinal study of the chunk use by Chinese college English majors. *Foreign Language World*, (5), 34-41.
- [15] Wen Qiufang. (2006). Tendencies and features of English majors' spoken vocabulary change. *Foreign Language Teaching and Research*, (3), 189-195.
- [16] Wolfe-Quintero K, Inagaki S & Kim H. (1998). Second language development in writing: Measures of fluency, Accuracy and

Complexity. Hawaii: University of Hawaii Press.

- [17] Wray A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- [18] Wray A. (2008). *Formulaic Language: Pushing the Boundaries*. Cambridge: Cambridge University Press.
- [19] Xu Jiajin & Liang Maocheng. (2010). *Using corpora: A practical coursebook*. Beijing: Foreign Language Teaching Research Press.
- [20] Xu Xianwen. (2010). Study of the chunk type in writing of the postgraduate English learners. *World of Foreign Languages*. (5), 42-47.
- [21] Zhang Jianqin. (2004). Comparative study of the vocabulary phrases between Chinese advanced, medium and primary English learners. *Foreign Language World* (1), 10-14.

**Liwei Zhu** was born in Jilin China in 1981. She received her M.A. in linguistics from Beihua University, China in 2006. She is currently a lecturer in the School of Foreign Languages, Tianjin Polytechnic University, Tianjin, China. Her research interests include second language acquisition and sociolinguistics.