A Study on the Use of Lexical Chunks by Chinese EFL Learners in Writing

Ling Shi[1] & Lei Wang[2]

[1] Hangzhou College of Commerce, Zhejiang Gongshang University, China

[2] Foreign Language Institute, Zhejiang Gongshang University, China

Correspondence: Ling Shi, The Foreign Language Department, Hangzhou College of Commerce, Zhejiang Gongshang University, Zhejiang, China. E-mail: shilinghelen@163.com

.

Abstract

Lexical Approach put forward by Michael Lewis (1993) is widely acknowledged in EFL teaching and lexical teaching is very important development in the evolution of language teaching (Lowe, 2003). For about thirty years of teaching English as a foreign language (EFL) in China, more and more teachers have realized the importance of teaching and encouraging learners to use ready-made lexical chunks. However, the present study focuses on the overuse of lexical chunks in learners' writings in a high stake national test (College English Test Band 6 – CET6). The corpus-based data analysis will be done to find the most commonly used lexical chunks by Chinese EFL learners and demonstrate what is meant to be the overuse of lexical chunks. Furthermore, the reasons for misuse and overuse of lexical chunks will be discussed. The findings drawn from structural and functional analysis of lexical chunks also have some pedagogical implications.

Keywords: lexical chunks, corpus-based, Chinese EFL Learners, writing

1. Introduction

It begins with an interesting story about overuse of lexical chunks. One of the raters for CET6 (College English Test Band 6), a college English teacher, wrote for fun a short essay following what we have found very popular among students' compositions—the piling of ready-made lexical chunks, mocking the overuse of some of the chunks in writing composition in CET6. The topic of writing for this year is "View on the University Ranking". Similar to the usual format, students are required to write in a three paragraph essay with the first paragraph being a general introduction of the popularity of university ranking, second stating both the cons and pros and third airing writer's own opinions.

Raters, after reading hundred and thousands of student work, were really fed up with some of the lexical chunks frequently used by students. For some writings, you may have the feeling that once you took all those lexical chunks away, nothing was left, not to mention expressing personal views on the topic given. Here comes the sample of writing written by a rater to mock the overuse of lexical chunks. The underlined part is frequently used lexical chunks and Relativity Index Ranking (RI) mentioned in the article refers to the standard to evaluate the quality of composition rating. If RI is too low, say less than 0.35 for CET6, the rate is not supposed to be qualified.

Sample

Nowadays, Relativity Index Ranking has <u>become a popular phenomenon</u>. RI Ranking <u>has gained increasing popularity</u> among teachers who are here in ZJU checking CET6 writing, especially those whose index is lower than 0.35.

Over this issue, <u>different teacher has different opinion</u> toward it. <u>Some are enthusiastic</u> about RI

Ranking; they argue that RI is <u>far more important than</u> the quantity of the writing you've checked. <u>While others think</u> quantity should <u>outweigh</u> RI. <u>Only we could</u> check at least 350 passages <u>can we</u> finish the task.

<u>As far as I am concerned</u>, <u>as the saying goes</u>, <u>every coin has two sides</u>, <u>both opinions are reasonable to some extent</u>. <u>Personally, I think</u> quality is <u>much more important</u>. <u>So in a word</u>, RI Ranking is more gelivable!

What score can we give to a composition like this? In writing, overused phrases and the so-call multi-purpose structure are heavily piled up, such as "as far as I am concerned", "as the saying goes", "every coin has two sides", etc. The above example raises a question in EFL teaching. While many previous studies in SLA emphasized on the importance and effectiveness of teaching and learning lexical chunks in EFL, it is doubtful whether the more the better. In this study, we will use Chinese EFL learners' writings as data to discuss what Chinese EFL learners' problems are in using lexical chunks, what the appropriate way to teach lexical chunks is and how to encourage communicative creative writing.

2. Theoretical Background

Formulaic language has in recent years become widely recognized as a crucial aspect of second language competence. Jespersen (1924) was the first to make a general distinction between "formulas" and "free expressions" which "pervades all parts of grammar." Bloomfield (1933) observed that "many forms lie on the border-line between bound forms and words, or between words and phrases" (p. 181). Firth (1957), developed the idea of *polysystematism*, which is famous for the quotation "you shall judge a word by the company it keeps".

Starting from the 1960's, Chomsky's approach to syntactic structure gained prominence. He considered syntactic competence permits grammatical strings or sentences to be generated word by word, but not all grammatical sentences can perform any functions, only certain of these syntactically correct strings or sentences are assigned particular functions in particular contexts. Some phrases and expressions have become conventionalized as more or less unanalyzed composites of form and function. Since 1980s there has been a growing trend in language acquisition research to recognize words as inextricable entwined with a range of syntagmatic contexts and contextual patterns as opposed to viewing them as discrete units that can be mastered in isolation (Pawley & Syder, 1983; Sinclair, 1991; Nattinger & DeCarrico, 1992; Howarth, 1998; Cowie, 1998; Ellis, 2001; Wray, 2002). At the same time, the use of Lexical Approach put forward by Michael is widely acknowledged in EFL teaching and lexical teaching is very important development in the evolution of language teaching (Lowe, 2003). The lexical approach concentrates on developing learners' proficiency with lexis, or words and word combinations. It is based on the idea that an important part of language acquisition is the ability to comprehend and produce lexical phrases as unanalyzed wholes, or "chunks," and that these chunks become the raw data by which learners perceive patterns of language traditionally thought of as grammar (Lewis, 1993, p. 95). Teachers give instruction focusing on relatively fixed expressions that occur frequently in spoken language, such as, "I'm sorry," "I didn't mean to make you jump," or "That will never happen to me," rather than one originally created sentences (Lewis, 2000). Michael Lewis (1993) also suggests that the key principle of a lexical approach is that "language consists of grammaticalized lexis, not lexicalized grammar." Lexical approach advocates argue that language consists of meaningful chunks that, when combined, produce continuous coherent text, and only a minority of spoken sentences are entirely novel creations.

However, there are researchers who started to question the lexical approach and remind us on the importance of both grammar and communicative goals in verbal communication. Lowe (2003) pointed out:

The lexical view of language is not 'the answer'—at least not the whole answer—because, even if we use Michael Lewis' taxonomy of phrase-types (fixed expressions, semi-fixed expressions, phrases, collocations, and words), and even if these phrase-types incorporate the majority of the words of the language, this does not account for how we put these together into communicative strings.

He exemplifies how a Chinese learner improved her writing by specific instruction a priority list of grammatical points on syntax. He emphasizes that some apparently complex sentences have simple underlying structures. For learners with European language as their mother tongue, such simple structures are easy to learn, But for Chinese learner, these is the place where simple lexical approach could not cover and need to be instructed with much more effort.

3. Methodology of the Present Study

With the overuse of lexical chunks in learners' writing, the fulfillment of the communicative goal is less

satisfactory. The present study firstly set up a small corpus (See Table 1) by collecting 48 pieces of CET 6 learners' composition which are typed out from the electronic scanning version. No correction of any kind is made in terms of grammatical mistakes. The choice of the sample writings is a random one, without any consideration of the scores given by the raters.

Table 1. Basic information of CLC established in the present research

Corpus	Size of corpus	No. of texts	Tokens	Type	TTR	Average length of text	Sources of texts
Chinese EFL Learner Corpus (CLC)	12,475	48	11,014	1,486	13.5%	229	Chinese EFL learners' composition for CET6, 2013

Secondly, the corpus-based data analysis has been carried out. We counted the most frequently used lexical chunks by using automated corpus tool Power Conc. 3.2 (See Figure 1 Extraction of Power Conc.3.2) to find out the lexical chunks favored by Chinese learners in their writings and demonstrated what we mean by the overuse of lexical chunks. In order to generate a list of refined lexical chunks, the key criterion of the length and frequency thresholds set in Power Conc. 3.2 is 4-word lexical chunks occurring 3 times or more. Four-word sequences are found to be the most researched length for writing studies, probably because the number of 4-word chunks is often within a manageable size for manual categorization and concordance checks (Chen & Baker, 2010).

Thirdly, the structural and functional analysis based on the list of refined lexical chunks generated by Power Conc. 3.2 has been conducted with some statistical analysis.
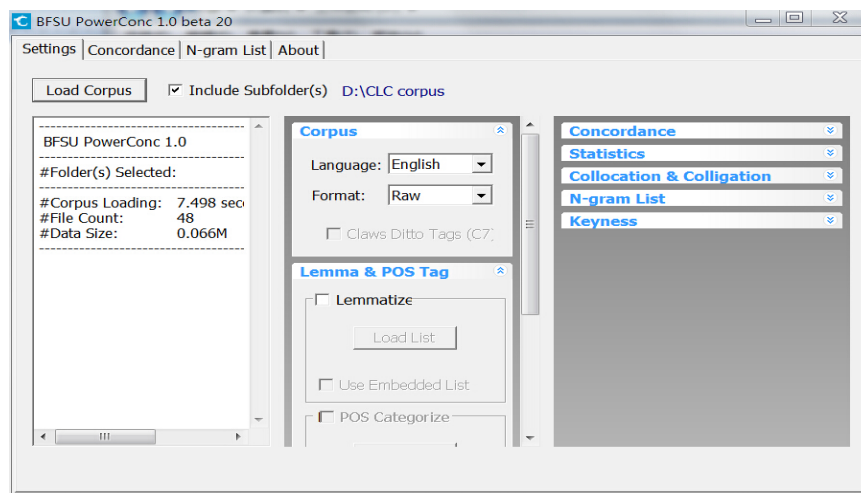

Figure 1. Extraction of Power Conc. 3.2

4. Findings and Discussion on Data Analysis

As can be seen in Table 2, the general size of corpus for the present research is 12,475 words, and the number of 4-word lexical bundles used by the Chinese EFL learner accounts for 20.45%, which takes a high percentage of the average length of an article. Table 3 shows the top 10 most frequent 4-word lexical chunks used by Chinese EFL learners in writing. It is observed that of those 10 top lexical chunks, 9 lexical chunks are all functional, while the rest one 'the university ranking is' is the writing topic related. Based on the four-word lexical bundles identified in CLC, the findings from this research show that Chinese EFL learners rely much on functional lexical chunks in writing, but their chunk diversity is quite limited. Such lexical chunks as "as far as", "on the other hand", "last but not least", ect. are overused in EFL learners' writing.

Table 2. The overall distribution of four-word lexical chunks in CLC

Corpus	Size of corpus	Number of texts	Frequency	No. of lexical chunks (Types)	No. of lexical chunks (Tokens)	TTR Type-token ratio
Chinese EFL Learner Corpus (CLC)	12,475	48	>3	64	313	20.45%

Table 3. Top 10 most frequent four-word lexical chunks in CLC

	Chinese EFL Learner Corpus (CLC)			
No.	Target lexical chunks	Freq	Distribution	Average occurrence ratio
1	As far as I	19	19	39.58%
2	The university ranking is	19	15	31.25%
3	On the other hand	14	14	29.17%
4	Different people have different	8	8	16.67%
5	Last but not least	8	8	16.67%
6	As I am concerned	7	7	14.58%
7	When it comes to	7	6	12.50%
8	Coin has two sides	6	6	12.50%
9	I agree with the	6	6	12.50%
10	With the development of	6	6	12.50%

Notes. Freq = frequency; Distribution= the number of text in which the target chunk occurs; average occurrence ratio: how many chances a target chunks occurs in a text on average.

4.1 Structural Analysis

When it comes to the structural characteristics, the NP-based and PP-based chunks in CLC reach 25.1% and 21.88% respectively (See Figure 2), indicating that Chinese EFL learners in this research have acquired the use of nominalization and prepositionalization in writing. The Adj.-based chunks only take 1.56%, the smallest proportion in CLC.
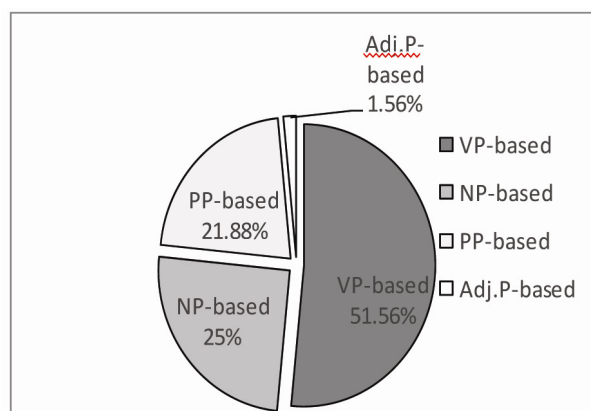

Figure 2. Proportional distributions of lexical chunks (types) in CLC

Meanwhile, EFL learners are more likely to use VP-based lexical chunks, which reach 51.56% and take the largest proportion in CLC. According to Table 4, it shows Chinese EFL learners' preference for copula-be and that-clause in writing, for example, *"(it) is bad for"*, *"most important thing is"*, *"hold the view that"*, *"some people think that"*, etc. "NP +copula be" chunks take the largest proportion at 20.31% in VP lexical bundles (See Figure 3), which indicates Chinese EFL learners rely much on the simple verb structure to express idea. It is perceived that though most of them had over 6 years' English learning, students are still weak at flexibly applying different communicative structure. Figure 3 also indicates Chinese EFL learners' reluctance of using "Passive verb +PP fragment" chunks in writing, which only take 1.56% in CLC.
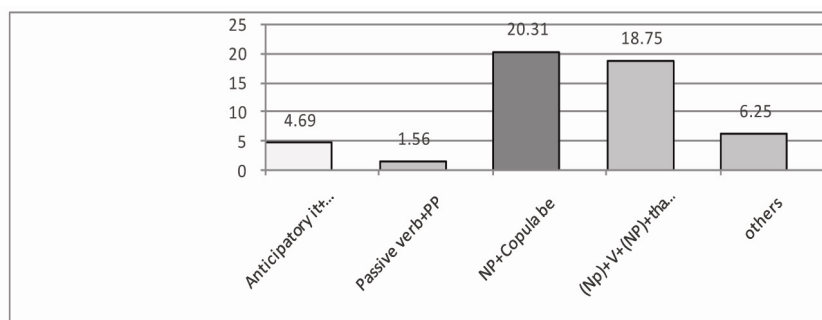
Figure 3. Proportional distributions of VP lexical chunks (types) in CLC

Table 4. Proportional distribution of lexical chunks (types) across the structural categories in CLC

Structural patterns		Chinese EFL Learner Corpus (CLC)		Examples
		Types of lexical chunks	Proportional Distribution	
NP-based	1.NP+post-modifier fragment	16	25.00%	attention to university ranking, a large number of, etc.
PP-based	2.PP + noun phrase fragment	14	21.88%	of the university ranking, on the one hand, with the development of, from person to person, etc.
VP-based	3.Anticipatory it + VP/ AP	3	4.69%	it is bad for, it is unfair for
	4.Passive verb + PP fragment	1	1.56%	as I am concerned
	5.NP+Copula be	13	20.31%	the university ranking is, most important thing is, etc
	6.(NP)+V+(NP)that-clause fragment	12	18.75%	hold the view that, people have different opinion, some people think that, etc.
	7.Others	4	6.25%	every coin has two (sides), get a good job
Adj.P-based	8.Adj.P	1	1.56%	more and more popular
Total		64	100%	

4.2 Functional Analysis

The functional categorization adopted in the present research follows the categories devised by Biber and colleagues (Biber & Barbieri, 2007; Biber et al., 2003, 2004), namely referential expression, stance bundles, and discourse organizers (See Table 5). In detail, the referential expression aims to specify a given attribute or condition, and can be sub-categorized into framing, quantifying, and place /time. Stance bundles are supposed to express a writer's evaluation of a proposition, and can be sub-categorized into epistemic, obligatory and ability. Discourse organizers are functioned to structure text, and can be sub-categorized into topic introduction, topic elaboration, inferential and identification /focusing.

According to Figure 4, discourse organizers rank as the largest category in CLC, having the extremely highest proportion at 48.44%, while referential expressions only take up 10.93%. This result indicates a significant difference between EFL Learners' writing and native writing in terms of functional distribution of lexical types. As for the referential expression, Chinese EFL learners seem to use certain referential deictic expressions, such as *in recent years*, and *all over/around the world,* which do not frequently used by native writers. More examples extracted from CLC are listed as follows (See Table 6), and those certain referential expressions suggest an "EFL Learners chunk" rather than a "native chunk".

Table 5. The functional categorization adopted

The Functional Categorization	Function	Sub-categories	Example
Referential expression	To specify a give attribute or condition	Framing	in terms of
		Quantifying	in a number of
		Place/time	at the same time
Stance bundles	To express a writer's evaluation of a proposition	Epistemic	are more likely to
		Obligatory	it is necessary to
		Ability	to be able to
Discourse organizers	To structure text	Topic introduction	last bur nor least
		Topic elaboration	on the other hand
		Inferential	as a result of
		Identification/focusing	one of the most

Table 6. Extracted examples of the referential expression most frequently used in CLC

- *...but also all over the world $_{(7)}$, there are more and*
- *The universities all round the world $_{(7)}$ to think about whether...*
- *In recent years $_{(6)}$, people began to...*
- *In recent years $_{(6)}$, college ranking has come...*
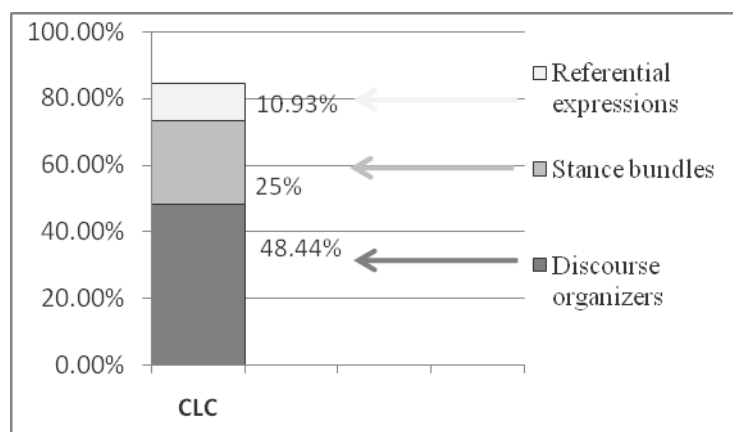

Figure 4. Functional distributions of lexical chunks (types)

As for the stance bundles, it shows that Chinese EFL learners employ the small range of epistemic chunks, and cannot flexibly employ hedging expressions. Based on the present data-analysis, the typical and frequently used stance bundle in CLC is only "It + adj. fragment" frame. The examples extracted from CLC are listed as follows (Table 7). By contrast, proficient native writers use the wide range of epistemic bundles and hedging devices, including modal verbs (would *have to be, would be difficult to*), hedging verbs (seems *to have been, is has been suggested*), and hedging nouns *(there is no evidence that)* to qualify their propositions (Chen &Baker 2010).

Table 7. Extracted examples of the stance bundles most frequently used in CLC

- *It is bad for $_{(3)}$ their healthy lifestyle.*
- *It is reasonable for $_{(3)}$ people to cast doubt*
- *Because they claim it is unfair for $_{(3)}$ the universities...*
- *It is important to $_{(2)}$ go on university ranking*

4.3 Discussion

It should be noted that the use of lexical chunks is complex and there is no one-to-one relationship between a chunk misuse and the reason of misuse. In general, there are two main reasons for misuse and a worse teaching-learning cycle for overuse, discussed as follows:

1) *Negative transfer in lexical bundles from Chinese at different levels:* Chinese EFL learners in this research, though as advanced English learners, are still influenced by their mother tongue in both grammar and thinking patterns. Grammatically, students would "play it safe" in chunk selection, and tend to place the common lexical chunks in a fixed way, for the adverbial expressions of time or places in Chinese cannot be moved within a sentence flexibly. In addition, they usually avoid certain lexical chunks which indicate relative clauses in English, such as *"the extent to which"*, *"the degree to which"*, etc., while these chunks are commonly used by English natives. With respect to the transfer of thinking patterns, Chinese EFL learners are deeply influenced by Chinese formulaic language and stylistic conventions in formal writing, such as *"with the development of"*, *"from the perspective of"*, *"pay more attention to"*, etc.

2) *The absence of pragmatic quality of lexical chunks used by Chinese EFL learners:* As is demonstrated from the present research data analysis, it reflects that Chinese EFL learners focus too much on the construction of the text itself while ignore the surrounding context. Meanwhile, they usually stick to the chunk instructions received in the early years of English learning and de-contextualize some lexical chunks in writing. In this research, Chinese EFL learners overuse the chunk *"it is necessary to"*, *"it is important that"*, etc. Though these are typical and widely-recognized structures in writing courses, they are too assertive and aggressive in writing practice.

3) *A worse teaching-learning cycle:* There are numerous reasons contribute to the poor situation, basically because of the test-oriented learning and teaching system. A number of researches suggest that many EFL textbooks often present unnatural and unrealistic dialogues which are not an accurate reflection of real world language use (Cheng &Warren 2007). For example, such functional lexical chunks as "I agree with you", "let's start our discussion" are frequently used to signal agreement, but unnecessary and only make the conversation unnatural. Moreover, both EFL teachers and students believe that using lexical chunks would effectively improve scores in the exam. Therefore, many EFL teachers prefer to teach writing sample model to students and overemphasize the importance of formulaic expression and lexical chunks in English courses.

5. Pedagogical Implications

This research on lexical chunks in Chinese EFL learner writing is significant in three aspects, EFL teaching, EFL learning, and improvement of EFL textbooks. First of all, it is conducive to EFL teaching that teachers can have a better knowledge of chunk distributions and uses through the identification and description of the most frequently-used four-word lexical chunks in the learners' corpora. Instead of emphasizing on typical lexical chunks, teachers are aware of three criteria in classroom instructions: 1. what to teach; 2. how much to teach; and 3. to what degree should classroom instructions reach. In this way, the overuse and underuse of lexical chunks can be avoided to some extent, which helps EFL learners achieve native-likeness in writing.

Secondly, this research is beneficial for EFL learners' lexical bundle acquisition. As Yorio (1989) claims: "Unlike children, adult L2 learners do not appear to make extensive early use of prefabricated, formulaic language, and when they do, they do not appear to be able to use it to further their grammatical development". Therefore, Chinese EFL learners should be cautious about language contexts and become aware of the lexical chunk usage with surrounding context, as classroom instructions and dictionary entries are usually de-contextualized in chunk explanations. In this way, Chinese EFL learners can keep a balance between the bundle productivity and quality, making their language more acceptable and native-like.

Finally, the findings in this research would contribute to the improvement of EFL textbooks. With the help of corpus-driven results, the compilation of textbooks can take into consideration the structural and functional distributions and characteristics of lexical chunk, as well as the common pragmatic failures in chunk usage by EFL learners. Textbooks with "real examples" identified through corpora retrieval would be more persuasive in instructing and meanwhile improve learners' understanding of chunk usages and selection. Such kind of textbooks, together with classroom instructions, would better meet learners' needs than traditional textbooks with word or lexical chunk lists.

6. Limitations of This Research

Though this paper has attempted to investigate the 4-word lexical bundles in Chinese EFL learners' writing, some limitations still exist in the research. First of all, the CLC corpus contains 12,475 running words, which is quite limited in size. A fuller account of lexical chunks or formulaic language entails the expansion in the size of corpus. Secondly, given the manageability of lexical chunk analysis, this research focuses only on the 4-word lexical chunks. Further discussions on various lengths of lexical chunks are also necessary in the future study. Thirdly, due to the limitation of retrieval tool, this research does not cover the incomplete structures of lexical phrases, such as "not only… but also…", "the …er, the…er", etc. It should be acknowledged that such incomplete structures also have functional status in various discourses and play an important role in writing.

Acknowledgements

References

Biber, D. (2009). A corpus-driven approach to formulaic language in English Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14*(3), 275-311. http://dx.doi.org/10.1075/ijcl.14.3.08bib

Biber, D., & Barbieri, F. (2007). Lexical Bundles in University spoken and written registers. *English for Specific Purposes*, *26*(3), 263-286. http://dx.doi.org/10.1016/j.esp.2006.08.003

Biber, D., Conrad, S., & Cortes, V. (2003). Lexical Bundles in speech and writing: An initial taxonomy. In A. Wilson, P. Rayson, & T. McEnery (Eds.), *Corpus Linguistics by the Lune* (pp. 71-93). Fankfurt/Main: Peter Lang.

Biber, D., Susan, C., & Viviana, C. (2004). If you look at Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, *25*(3), 371-380. http://dx.doi.org/10.1093/applin/25.3.371

Chen, Y. H., & Baker, P. (2010). Lexical Bundles in L1 and L2 Academic Writing. *Language Learning & Technology*, *14*(6), 30-49.

Cowie, A. P. (1998). *Phraseology: Theory, Analysis and Applications* (pp. 1-20). Oxford: Oxford University Press. http://dx.doi.org/10.2307/417696

Ellis, N. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and Second Language Instruction*. Cambridge: Cambridge University Press.

Firth, J. R. (1959). *Papers in Linguistics*. London: Oxford University Press.

Howarth, P. (1998). The Phraseology of Learners' Academic Writing. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 161-186). Oxford: Oxford University Press.

Jespersen, O. (1924). *The philosophy of grammar*. London: George Allen & Unwin. http://dx.doi.org/10.4324/9780203716045

Lewis, M. (1993). *The Lexical Approach: The State of ELT and the Way Forward*. Hove: Language Teaching Publications. http://dx.doi.org/10.2307/3587574

Lewis, M. (2000). Learning in the lexical approach. In M. Lewis (Ed.), *Teaching collocation. Further developments in the lexical approach* (pp. 155-185). Hove: Language Teaching Publications.

Lowe, C. (2003). Lexical Approeachs now: the role of syntax and grammar. *IH Journal of Education and Development*.

Nattinger, J., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. Oxford: Oxford University Press. http://dx.doi.org/10.2307/329018

Pawley & Syder. (1983). Two puzzles for linguistic theory: Native-like selection and native-like fluency. In J. Richards & R. Schmidt (Eds.), *Language and Communication* (pp. 191-226). London: Longman.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press. http://dx.doi.org/10.2307/330144

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/cbo9780511519772

Yorio, C. (1989). Idiomaticity as an Indicator of Second Language Proficiency. In K. Hyltenstam & L. Obler

(Eds.), *Bilingualism across the Lifespan* (pp. 55-72). Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/cbo9780511611780.005

Appendix A

The 50 Most Frequent Four-word Lexical Chunks in CLC

CLC Corpus Size	12,475		Hit(s)	7,162		Item(s)	6,739
No.	**4-word Lexical Chunk**	**Freq.**	**No.**		**4-word Lexical Chunk**		**Freq.**
1	As far as I	19	26		a lot of attention		3
2	the university ranking is	19	27		but not the least		3
3	On the other hand	14	28		by the university ranking		3
4	far as I am	9	29		development of the society		3
5	Different people have different	8	30		for us to choose		3
6	Last but not least	8	31		From my point of		3
7	as I am concerned	7	32		from person to person		3
8	when it comes to	7	33		I think University Ranking		3
9	coin has two sides	6	34		in this way can		3
10	I agree with the	6	35		It is bad for		3
11	Some people hold the	6	36		it is unfair for		3
12	with the development of	6	37		Last but not the		3
13	Every coin has two	5	38		more and more popular		3
14	On the one hand	5	39		most important thing is		3
15	Only in this way	5	40		my point of view		3
16	people have different opinions	5	41		one of the most		3
17	attention to university ranking	4	42		People have different opinion		3
18	get a good job	4	43		people hold the view		3
19	hold the view that	4	44		Some people think it		3
20	it is a good	4	45		some people think that		3
21	more and more people	4	46		the high school students		3
22	people think it is	4	47		the most important thing		3
23	university ranking has been	4	48		the rapid development of		3
24	university ranking is very	4	49		the university ranking may		3
25	A large number of	3	50		think it is a		3

Copyrights