On Metadiscoursal Features of Chinese University Students' Oral English: A Perspective From Chunks[*]

LIN Wei-yan

Jinan University, Guangzhou, China

This paper explores metadiscoursal features of Chinese university students' oral English based on their chunks in the hope of providing implications for oral English teaching in Chinese universities. Data consisted of oral English produced by university students from China and Britain. Form, meaning, and functional features of high frequency four-word OCs (oral chunks) with metadiscoursal features (i.e., metadiscoursal oral chunks, or MOCs) used by university students from the two countries were compared. It is found that similar to British university students' MOCs, Chinese university students' MOCs share salient lexical-grammatical patterns and signal interpersonal, textual and mixed functions which are highly related to their patterns. However, Chinese university students' MOCs are significantly underused in both types and tokens, formally less diversified and more dysfluent, and functionally more assertive, suggesting the students' poorer metadiscoursal performance. The results imply that OCs can be utilized to evaluate EFL (English as a foreign language) learners' metadiscoursal performance.

Keywords: metadiscourse, chunk, corpus, EFL learner

Introduction

As stated by Sinclair and Renouf (1988, p. 156), language text is simultaneously organized on at least two different dimensions, or "planes". The first plane is used for constructing and elaborating messages (propositional meanings), and the second is concerned with commenting on, labeling, evaluating, and generally negotiating the messages interpersonally. However, the importance of purposes, intentions, and objectives of language users has not, until recently, been recognized and the second of these planes has been largely neglected. The second plane of language consists of the topic of this paper—metadiscourse. This paper aims to investigate metadiscoursal features of Chinese university students' oral English from the perspective of chunks.

As a "central pragmatic construct" (Hyland, 1998, p. 437), metadiscourse can be understood as "self-reflective linguistic material referring to the evolving text and to the writer and imagined reader of that text" (Hyland & Tse, 2004, p. 156). This paper focuses on metadiscourse in interactional oral English; thus, by making reference mainly to William (1981, pp. 211-212) and Hyland (1998), metadiscourse is defined as the linguistic material in discourse that does not add anything to the propositional content but that signals discourse

---

LIN Wei-yan, lecturer, Ph.D., College of Foreign Studies, Jinan University.

organization, interaction with listener or speaker stance towards either its content or the listener. Existing research has tended to focus on written metadiscourse, centering upon its influence on comprehension (Crismore, 1989; Crismore & Vande Kopple, 1997), its cross-cultural differences (Crismore, et al., 1993; Dahl, 2004), effects of its explicit teaching in writing (CHENG & Steffensen, 1997; Shaw & LIU, 1998), patterned developmental changes EFL (English as a foreign language) student writers experience across four years (H. M. XU, 2001), etc.. However, the association between the use of metadiscourse and EFL learners' pragmatic performance still remains under-researched. Since metadiscourse belongs to the ubiquitous and yet neglected second plane of language, this study on EFL learners' metadiscoursal performance may provide implications for EFL teaching.

Metadiscourse can be in the forms of single words (*but*, *yeah*), phrases (*by the way*, *on the other hand*), or clauses (*I think*, *as I said*). This paper regards OCs (oral chunks) as the analytical units in light of the "phraseological tendency" of language, "where words tend to go together and make meanings by their combinations" (Sinclair, 2004, p. 29) . OCs are defined here as continuous recurrent sequences of at least two words automatically retrieved from an oral corpus. Simply put, automatic retrieval of OCs is to run certain software tool to split and retrieve continuous four-word sequences at a certain frequency cut-off. Let us suppose that an oral text begins like this: "Once upon a time there was a beautiful princess". If we run WordSmith 5.0 (Scott, 2008) and choose clusters with the length of four words at a frequency cut-off of one (i.e., each cluster appears at least once in the corpus), the following six OCs will be retrieved: *once upon a time*, *upon a time there*, *a time there was*, *time there was a*, *there was a beautiful*, and *was a beautiful princess*. Automatic retrieval necessarily means that the software tool retrieves OCs which in most cases lack any syntactic or semantic integrity, as well as strings that display integrity of one or both kinds (McCarthy & Carter, 2002). Actually, a majority of the OCs are syntactic fragments, not constituting complete syntactic elements at phrasal or clausal levels and are thus not recognized as fixed expressions by native speakers (see Altenberg, 1998; Biber, Johansson, Leech, & Finegan, 1999; McCarthy & Carter, 2002; WEI, 2007). Consequently, "[c]onventional grammars would certainly dismiss these as incomplete structures" (O'Keeffe, McCarthy, & Carter, 2007, p. 70). Nevertheless, results of previous studies have shown that high frequency OCs "turn out to be readily interpretable in both structural and functional terms" and thus "should be regarded as a basic linguistic construct with important functions for the construction of discourse" (Biber, Conrad, & Cortes, 2004, p. 398-399). Besides, high frequency OCs tend to have metadiscoursal, pragmatic functions (Altenberg, 1998; Biber et al., 2004; O'Keeffe et al., 2007). However, existing research on EFL learners' OCs has hardly focused on their metadiscoursal features. Therefore, this study attempts to interpret Chinese university students' English metadiscoursal performance, especially in the aspect of utterance initiation, through an analysis of their MOCs (metadiscoursal oral chunks) in the hope of providing implications for oral English teaching in Chinese universities.

Research Design

The main research question is: What are the features of MOCs used by Chinese university students in comparison with those by British university students in terms of general frequencies, linguistic patterns, and functions?

Research data consisted of two corpora, as displayed in Table 1.

Table 1

Research Data

	Corpora	Subjects	Size (tokens)
Target	LINDSEI-CH (B turns)[①]	53 Chinese university students	63,463
Reference	LOCNEC (B turns)	50 native British university students	117,703

Note.[①] In the corpora, B turns consist of the subjects' (i.e., interviewees') speeches, while A turns consist of those of the teachers (i.e., interviewers).

LINDSEI-CH is a sub-corpus of LINDSEI (Louvain International Database of Spoken English Interlanguage), which is made up of informal interviews divided into three parts: set-topic talk, free discussion, and picture description between university undergraduates in English (usually in their third or fourth year) in their twenties and teachers (Gilquin, De Cock, & Granger, 2010a). LOCNEC (Louvain Corpus of Native English Conversation) was compiled as a reference corpus of native British university students, built according to the same principles as LINDSEI (Gilquin, De Cock & Granger, 2010b, p. 65). Thus, the two sets of data are highly homogeneous in terms of tasks, context, and interview durations.

For the retrieval of OCs, this research employed WordSmith's (Scott, 2008) function WordList cluster.

In light of Hyland's (1998) definition of metadiscourse, MOCs are defined as OCs of four contiguous words that can be used to initiate utterances and consist of non-propositional elements. Accordingly, MOCs should meet three requirements: utterance-initiating signals, non-propositional elements, and four contiguous words (LIN & HE, 2012).

In terms of research methodology and procedure, typical linguistic patterns of British university students' MOCs (NS MOCs) were first constructed from nearly one thousand four-word OCs by implementing the working procedure of constructing meaningful units proposed by Lexical Grammar (Sinclair, 2004) in corpus linguistics. They then served as reference to reveal features of Chinese university students' MOCs (CH MOCs) in terms of general frequencies, linguistic patterns, and functions. Functions and sub-functions of MOCs were analyzed mainly within their four-word context. This functional taxonomy was mainly based on Hyland's (2005, pp. 48-54) classification of written metadiscourse in English, which was complemented by Erman and Warren's (2000, pp. 43-45) classification of prefabs (i.e., prefabricated chunks) in spoken English. Three broad functional categories of MOCs were finally distinguished: textual, interpersonal, and mixed-functional. Textual MOCs signal organization of a discourse (*that's to say I, in the first place*). Interpersonal MOCs signal interaction with the listener or speaker stance towards the listener or discourse content (*you know it was, I couldn't believe it*). Mixed-functional MOCs signal both discourse organization and listener interaction/speaker stance; that is, they have both textual and interpersonal functions (*but I don't know, and it was sort of*). Sub-functions of each broad category were also identified.

Results

Shared Features of Chinese and British University Students' MOCs

First, in terms of forms and meanings, they share two typical patterns: (1) "*and/ERM/YEAH/but/so +* pronoun-type subject" (*erm I don't know, but it was a*); and (2) "*I/it/that/you +* stance-type words"[1] (*I think it*

---

[1] In the patterns, slashes (/) mean "or" and capitalized words in italics include their different spellings in oral English. For example, *ERM* consists of *erm*, *er*, *eh*, and *mm*.

was, it was sort of). Pattern (1) is a formal pattern consisting of OC-initial words and their following grammatical elements. It is similar to colligation in Lexical Grammar (Sinclair, 2004, p. 164) in that it displays a grammatical choice (i.e., pronoun-type subject) colligating OC-initial words (i.e., conjunctions *and/so/but* or particles *ERM/YEAH*). Pattern (2) is a form-meaning pattern consisting of OC-initial words and their following words with certain semantic meaning, or stance. It is similar to semantic preference in Lexical Grammar in that it displays high frequency OC-initial words (i.e., pronouns *I/it/that/you*) co-selected with stance-type words (e.g., *think*, *mean*, *know*, and *sure*).

Second, in terms of functions, MOCs in both CH and NS signal interpersonal, textual, and mixed functions, with interpersonal MOCs taking up the highest proportions (see Table 2). A more detailed analysis has revealed that hedges are of the biggest percentage of all interpersonal MOCs in both CH and NS.

Table 2

General Functions of MOCs in CH and NS (Tokens)

Functions	CH		NS	
	R. Freq.[①]	N. Freq.[②]	R. Freq.	N. Freq.
Interpersonal	568 (49.1%)	895	1,797 (55.9%)	1,527
Mixed-functional	386 (33.4%)	608	761 (23.7%)	647
Textual	203 (17.5%)	320	654 (20.4%)	556
Total	1,157 (100%)	1,823	3,212 (100%)	2,729

Notes. [①] R. Freq. = Raw frequency; [②] N. Freq. = Normalized frequency (per 10,000 words).

Third, a strong relationship between MOC patterns and functions is found in both CH and NS, as displayed in Figure 1.
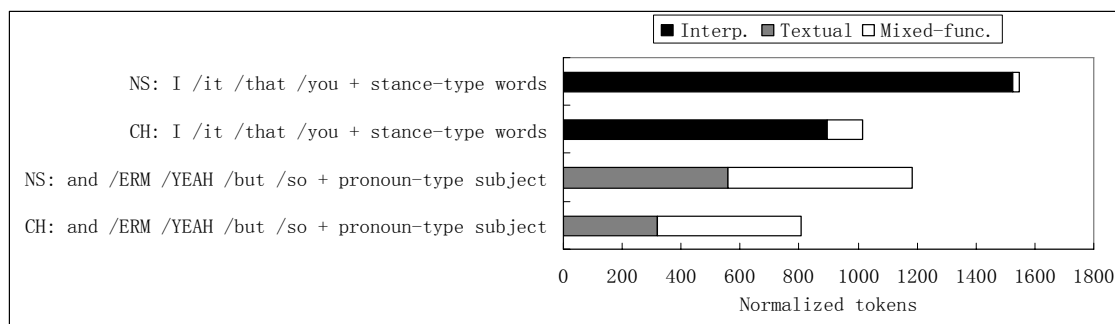

Figure 1. Relationship between linguistic patterns and functions of MOCs in CH and NS.

Distinct Features of Chinese University Students' MOCs

Underuse in frequency. MOCs are significantly underused in CH in both types and tokens, as displayed in Table 3.

This suggests that CH OCs have a weaker metadiscoursal tendency. This finding corresponds to what was found in previous studies on pragmatic features of Chinese EFL learners' OCs. For example, HE and M. F. XU (2003) found that EFL learners of four different L1 backgrounds (Chinese, Japanese, French, and Italian) used fewer two-word English "small words" (i.e., discourse markers that are highly frequent in speech, such as *well* and *sort of*) than English native speakers. DENG (2007) found that Chinese EFL learners significantly underused interactional formulaic sequences (*that's a really good point, but I think, yeah definitely*). Finally, WEI (2007)

found that Chinese EFL learners underused English OCs which basically perform pragmatic functions, such as epistemic tags (*I don't know*) and vagueness expressions (*sort of*).

Table 3

Frequencies of MOCs in CH and NS

	Types					Tokens				
	CH		NS		x^2		CH		NS		x^2
	R. Freq.	N. Freq.	R. Freq.	N. Freq.			R. Freq.	N. Freq.	R. Freq.	N. Freq.	
OC	2,047	3,226	3,642	3,094	+2.29[①]		5,346	8,424	9,570	8,130	+4.65
MOC	415	654	1,118	950	-42.68***[②]		1,157	1,823	3,212	2,729	143.36***
MOC %	20.3		30.7				21.6		33.6		

Notes. [①] "+" indicates overuse in CH; [②] "-" indicates underuse in CH; *** indicates significant difference at 0.001 level.

The weaker metadiscoursal tendency of CH OCs is best evidenced by the small percentage of their MOCs among the top 15 OCs in comparison with those in NS, as listed in Table 4.

Table 4

Top 15 OCs in CH and NS

Order	CH		NS	
	OCs	N. Freq.	OCs	N. Freq.
1	*to draw a picture[①]	41	it was it was	37
2	to be a teacher	32	**I don't know I**	32
3	*look at the picture	30	and things like that[③]	28
4	*draw a picture for	28	**erm I don't know**	20
5	like to talk about	28	at the end of	20
6	**I would like to**[②]	27	the end of the	20
7	I want to be	25	**I think it was**	16
8	*a picture for her	25	**I'd like to go**	16
9	at that time I	24	a bit of a	16
10	*draw a picture of	22	**it was really good**	14
11	*in the picture is	22	and stuff like that	14
12	in the in the	20	or something like that	14
13	*woman in the picture	20	**I don't know if**	13
14	*a picture of her	19	**I think I think**	13
15	how to say that	19	a lot of people	12

Notes. [①] OCs preceded by * in CH are propositional OCs induced by task; [②] OCs in bold in both CH and NS are MOCs; [③] OCs underlined in NS are other OCs with specialized metadiscoursal functions.

As displayed in Table 4, in CH, only one OC is identified as MOC; in NS, however, seven OCs are MOCs. In addition, apart from the seven MOCs in NS, four others signal specialized metadiscoursal functions (e.g., hedging) although they do not signal utterance initiation (which is one requirement of MOCs). The two groups of OCs take up 73.3% of all 15 OCs, making salient the non-propositional tendency of high frequency NS OCs. In contrast, CH OCs mainly indicate propositional meanings. As displayed in Table 4, eight CH OCs consist of *picture* and/or *woman*, which were induced by the content of the interview task during which the students were presented pictures showing an artist drawing pictures for a woman. Students in both CH and NS needed to finish the same task, and yet the proportions of their MOCs are so different.

Lower diversity and unusual dysfluency in form. The top 15 MOCs are much less diversified in CH than in NS, as listed in Table 5.

Table 5

Top 15 MOCs in CH and NS

CH		NS	
MOCs	N. Freq.	MOCs	N. Freq.
I would like to	27	I don't know I	32
*I think I will[①]	17	erm I don't know	20
*I think er the	16	I'd like to go	16
I'd like to talk	16	*I think it was	16
and I want to[②]	14	it was really good	14
er I mean the	14	I don't know if	13
er I want to	14	*I think I think	13
so I want to	14	and it was really	11
I don't know how	14	I thought it was	11
*I I think I	14	you know what I	10
*I think it is	13	and it was like	10
*so I I think	11	and it was just	9
I just want to	11	it was a bit	9
I will go to	11	it was just like	9
*er I I think	9	I was going to	9

Notes. [①] MOCs preceded by * consist of *I think*; [②] MOCs underlined consist of *I want to*.

In CH, six MOCs consist of *I think* (as preceded by *) and whereas in NS, only two consist of that expression. This finding is consistent with what was found in HE and M. F. XU (2003), i.e., *I think* was significantly overused by Chinese advanced EFL students and it was actually more frequent than 18 other common "small words" (discourse markers such as *ah, just, like, a bit, all right*). Similarly, J. J. XU and Z. R. XU (2007) found that Chinese university students overused *I think*. Besides, four CH MOCs consist of *I * want to* (as underlined) and yet none consist of that expression in NS. Analysis of the immediate right collocates of *I * want to* in the CH corpus at a frequency cut-off of five searched by WordSmith's (Scott, 2008) function Concord has revealed that *I * want to* tend to be followed by *BE* (18)[2], *GO* (15), *ERM* (6), and *TALK* (6). Among them, *TALK* was induced by one task of the interview during which the students were asked to choose one of the three topics to talk about. The three topics were: an experience you have had which has taught you an important lesson, a country you have visited which has impressed you, and a film/play you have seen which you thought was particularly good/bad (Gilquin et al., 2010b, p. 8). Induced by this task, CH students' topic choosing signals are highly stereotypical. Specifically, among the 53 CH students, 19 use *I... talk (about)...* (see Example (1)) and another 19 use *I... choose...*

Example (1)

okay em. I'll talk about one of the films that I think is a good one

em I'm going to talk about my trip to shao guan

eh I would like to talk about the first topic about the experience

---

[2] In this study, capitalized verbs are lemmas and the following figures in parentheses are their tokens in the corpus under research. For example, "*BE* (18)" indicates that all the various forms of *BE* (*am, are, is, was*) have a total token of 18.

eh I'd like to talk about a film I saw two weeks ago

em I like to talk erm topic the the second topic

em I would like to talk about a film I just seen. several days ago

er I'd like to talk my visit. to Beijing.

I'd like to talk about the place er one place

I'd like to talk about the topic three

mm okay I'd like to talk about an experience

ok. mm I'd like to talk about a film

okay I would like to talk about a place er im im impressed me very much

okay I'd like to talk about the most impressive film I've ever seen

okay okay er I would like to talk about er the film I like most

okay I try to talk about my experience

actually I just want to talked about erm

er I want to talk the place which give me. some very good impression

okay I want to talk about some. mm the life in the. when I came.

In comparison to Example (1), the topic choosing signals by NS students are much more diversified. Among the 50 students, 21 simply utter topic numbers (*er topic two*), 11 say out specific topics (*the country I visited*), and others either specify exact content of the topics chosen (*mhm okay well this summer I went over to America on holiday*), begin with *I... talk (about)* (*erm I'd like to talk about well a film and a play*), or start with *I... choose...* (*I think I'll choose the first topic I think*).

The unusual dysfluency of CH MOCs is evidenced by the significant overuse of their MOCs initiated by *ERM*. Table 6 compares sub-categories of "*and/ERM/YEAH/but/so* + pronoun-type subject" MOCs:

Table 6

Sub-categories of "and/ERM/YEAH/but/so + pronoun-type subject" MOCs in CH and NS

	CH		NS		x^2
Sub-categories	R. Freq.	N. Freq.	R. Freq.	N. Freq.	
ERM + pronoun-type subject MOCs	213 (41.5%)	336	228 (16.4%)	194	+33.62***[1]
and + pronoun-type subject MOCs	134 (26.1%)	211	566 (40.7%)	481	-77.24***[2]
so + pronoun-type subject MOCs	92 (17.9%)	145	186 (13.4%)	158	-0.38
but + pronoun-type subject MOCs	41 (8.0%)	65	201 (14.5%)	171	-34.04***
YEAH + pronoun-type subject MOCs	33 (6.4%)	52	210 (15.1%)	178	-48.25***
Total	513 (100%)	808	1,391 (100%)	1,182	-54.94***

Notes. [1] "+" indicates overuse in CH, *** indicates significant difference at 0.001 level; [2] "-" indicates underuse in CH.

As can be seen in Table 6, CH MOCs initiated with verbal fillers (*ERM*) form the highest proportion (41.5%); in contrast, such MOCs form a much smaller part in NS (16.4%). Actually, *ERM*-initiated MOCs are overused in CH while all other sub-categories are underused. Since verbal fillers are generally used for pauses when speakers are faced with production difficulty, their unusually high proportion in CH suggests that they may be regarded as unusual dysfluency. This result suggests that even at advanced level of English learning, Chinese university students' oral English is not fluent enough, and this rate of dysfluency may influence their discourse management and stance expression.

A more assertive tone in function. The assertive tone is evidenced by their significant underuse of hedges (i.e., a sub-category of interpersonal MOCs) in CH (see Table 7). Too assertive a tone may affect the

appropriateness of their oral English.

Table 7

Hedging MOCs in CH and NS

CH		NS		x^2
R. Freq.	N. Freq.	R. Freq.	N. Freq.	
297	468	997	847	-83.01***[1]

Note. [1] "-" indicates underuse in CH, *** indicates significant difference at 0.001 level.

Conclusions and Implication

Through an investigation of the metadiscoursal features of Chinese university students' oral English, this paper has found that CH MOCs are underused in both types and tokens, less diversified and more dysfluent in form, and more assertive in function. This result may suggest the neglect of non-propositional discourse in English teaching in China.

As an implication, OCs can be utilized to evaluate EFL learners' metadiscoursal performance. OCs are short, and yet they reflect certain features of EFL learners' oral English, such as lower diversity, unusual dysfluency, and high assertiveness. Those features may be found through other methods, but this method of analyzing OCs automatically retrieved from a corpus may be more objective and faster. It is objective since it can produce high frequency OCs that are "not available through intuition" (Read & Nation, 2004, pp. 30-31), and it is fast in that EFL learners' metadiscoursal performance can be quickly evaluated merely based on a small part of their oral English (i.e., high frequency OCs). Nevertheless, some issues remain to be discussed in the actual operation, including the setting of evaluation standards and complementary evaluation methods.

References

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A. Cowie (Ed.), *Phraseology: Theory, analysis and applications* (pp. 101-122). Oxford: Oxford University Press.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., Johansson, S., Leech, G., & Finegan, E. (1999). *The longman grammar of spoken and written English.* London: Longman.

CHENG, X. G., & Steffensen, M. S. (1997). Meta-discourse: A technique for improving student writing. *Research in the Teaching of English, 30*(2), 149-181.

Crismore, A. (1989). *Talking with readers: Meta-discourse as rhetorical act.* New York: Peter Lang.

Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written Communication, 10*(1), 39-71.

Crismore, A., & Vande Kopple, W. (1997). The effects of hedges and gender on the attitudes of readers in the United States towards material in a science textbook. In A. Duszak (Ed.), *Culture in style in academic discourse* (pp. 223-253). Berlin: de Gruyter.

Dahl, T. (2004). Textual metadiscourse in research articles: A marker of national culture or of academic discipline?. *Journal of Pragmatics, 36*(10), 1807-1825.

DENG, Y. C. (2007). *An investigation into formulaic sequences in Chinese EFL learners' spoken English*. Shanghai: Shanghai Jiaotong University.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20*(1), 29-62.

Gilquin, G., De Cock, S., & Granger, S. (2010a). *LINDSEI (Louvain international database of spoken English interlanguage)*. Retrieved from http://pul.uclouvain.be/fr/livre/?GCOI=29303100510890

Gilquin, G., De Cock, S., & Granger, S. (2010b). *LINDSEI (Louvain international database of spoken English interlanguage)*. Louvain: Presses universitaires de Louvain.

HE, A. P., & XU, M. F. (2003). Small words in Chinese EFL learners' spoken English. *Foreign Language Teaching and Research,* (6), 446-452.

Hyland, K. (1998). Persuasion and context: The pragmatics of academic discourse. *Journal of Pragmatics, 30*(4), 437-455.

Hyland, K. (2005). *Metadiscourse*. London and New York: Continuum.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*(2), 156-177.

LIN, W. Y., & HE, A. P. (2012). On metadiscoursal features of oral English "machine-cut" chunks. *Foreign Languages and their Teaching, 4,* 11-15.

McCarthy, M. J., & Carter, R. A. (2002). This, that and the other: Multi-word clusters in spoken English as visible patterns of interaction. *Teanga, 21,* 30-52.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.

Read, J., & Nation, P. (2004). Measurement of formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 23-35*)*. Amsterdam: John Benjamins.

Scott, M. R. (2008). *WordSmith Tools 5.0*. Liverpool: Lexical Analysis Software.

Shaw, P., & LIU, E. T. K. (1998). What develops in the development of second language writing. *Applied Linguistics, 19*(2), 225-254.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London and New York: Routledge.

Sinclair, J., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching*. London: Longman.

WEI, N. X. (2007). Phraseological characteristics of Chinese learners' spoken English: Evidence of lexical chunks from COLSEC. *Modern Foreign Languages,* (3), 280-328.

William, M. J. (1981). *Style: Ten lessons in clarity and grace*. Boston: Scott Foresman.

XU, H. M. (2001). *Use of meta-discourse markers by Chinese learners of English at the tertiary level: A developmental perspective*. Nanjing: Nanjing University.

XU, J. J., & XU, Z. R. (2007). Discourse Management Chunks in Chinese college learners' English speech: A spoken corpus-based study. *Foreign Language Teaching and Research,* (6), 437-443.