FORMULAIC LANGUAGE USE IN L2 CHINESE: THE ROLE OF PRE-WRITING PLANNING

By

Fei Fei

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2015

ABSTRACT

FORMULAIC LANGUAGE USE IN L2 CHINESE:
THE ROLE OF PRE-WRITING PLANNING

By

Fei Fei

The current study is framed within the SLA theories of formulaic language acquisition and task planning. The key question to be answered is whether directing learners' attention to the target Chinese formulaic sequences (FSs) during the pre-writing planning stage will prompt students to use such vocabulary more accurately, and consequently affect their written performance in terms of complexity, accuracy and fluency (CAF) measures. Given a mixture of results in previous studies concerning the role of pretask planning on complexity, accuracy and fluency both in oral and written language production, the goal is to find out which task conditions or task instruction contributes to variations in Chinese as a second language (L2 Chinese) learners' written language production.

Following Sangarun (2001, 2005) and Rott (2009), the study compared different pretask planning conditions through a quasi-experimental design based on their particular foci: content focused planning (CFP), language focused planning (LFP), content and language focused planning (CLFP), and planning with minimal guidance. The four pretask planning conditions were manipulated through the use of different planning sheets with pre-designed guidance on language forms, content or both. A total number of 67 CFL learners from an intensive study-abroad program in Beijing, China initially participated in the study. Participants' written outputs were analyzed using a battery of

CAF measures. Another three separate measures captured the target FS use, namely, the total number of target FSs per essay, non-target FS use and the ratio between correct use of FSs and total number of FSs.

Results indicated that planning instructions with an explicit guidance on linguistic forms (LFP) and on both forms and content (CLFP) were both more effective than other types of planning in promoting the use of the target FSs. The proportion of accurately used FSs and combined target and non-target FS use were positively related to one of the complexity measure (clauses per TTCU). Moreover, it was found that there were no trade-offs between accuracy and complexity measures, suggesting that learners from the CLFP group were likely to attend to both form and meaning in pretask planning.

Finally, the results suggest teaching and learning L2 Chinese formulaic language through writing instruction with carefully designed writing prompts and pre-writing guidance. It is likely that planning with a focus on form will prompt learners to use more target FSs and use them more accurately which promotes L2 learners' language development in complexity, accuracy and fluency.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my advisors and committee chairs, Dr. Debra Hardison and Dr. Xiaoshi Li, for their constructive comments, practical assistance, and endless patience throughout years. This dissertation would have never been possible without their guidance, encouragement and support.

My heartfelt thanks go to Dr. Aline Godfroid for her guidance on my thesis proposal, comments on the research design and statistical analyses. Her enthusiasm as a scholar is my constant inspiration, and her meticulous attitude towards details have accompanied me all the way to where I am today. I thank Dr. Patti Spinner for serving as a committee member and taking the time to read and comment on my work.

I am enormously indebted to my friend and colleague, Dr. Yun Lu, at Peking University for her assistance throughout the data collection process, and sharing with me her years of experience teaching Chinese as a second/foreign language.

I am very grateful to my mentor, Dr. Annela Teemant at the department of Education at Indiana University Purdue University – Indianapolis for her continuous help and support.

I would also like to extend my thanks and appreciation to my friends, Dr. Robin Fowler and Ms. Marilyn Schlief, for their encouragement, sincere friendship, and support. They read and commented on several drafts of my dissertation.

Finally, I am fortunate to have support and encouragement from my family. They truly deserve my greatest thanks and gratitude for their love and support which enabled me to complete this work.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Statement of the problem

There is a growing awareness that a large part of the discourse that we produce is composed of formulaic language (Wray, 2002, 2008). A mixture of terms has been attached to the definition of formulaic language, including “lexical phrases”, ”multi-word units”, “formulas”, “prefabricated chunks”, “ready-made utterances”, and so forth (e.g. Wray, 1999, p. 214). Formulaic language can range from completely free syntactic constructions (e.g., *NP be-TENSE sorry to keep-TENSE you waiting*) to completely fixed or frozen expressions such as *by and large* (Howarth, 1998, pp. 42-55). It is hypothesized that ultimate fluency in L2 is achieved by “clause-chaining” or stringing together a sequence of relatively independent clauses (Pawley & Syder, 1983, pp. 202-203) due to the limited memory space of human beings.

Formulaic language use has recently been found to be one of the strongest discriminators between lower and higher levels of second language (L2) English proficiency (Laufer & Waldman, 2011, regarding verb-noun collocation in L2 writing; Verspoor, Schmid, & Xu, 2012, regarding lexical chunks in L2 writing). However, the effects of teaching and learning of L2 formulaic language in the classroom setting are not always guaranteed (Lewis, 2000). It seems that learners only “learn what they are ready for (formulaic language) and in ways that may or may not match what the teacher does” (Lewis, 2000, p.2).

The Lexical Approach towards L2 vocabulary acquisition was proposed by Sinclair and Renouf (1998) and Lewis (1993, 1997, 2000) who put the acquisition of

collocations at the center of the lexical approach. Regarding L2 vocabulary learning and teaching, Hulstijn and Laufer (2001) tested the Involvement Load Hypothesis by comparing L2 English learners' initial learning and subsequent retention of new target words by learners from Israel and the Netherlands. They developed three tasks that represented three different levels of involvement: reading comprehension with marginal glosses, reading comprehension plus gap-fill, and writing a composition by incorporating the target words. Hulstijn and Laufer hypothesized that the writing task (with the target words) would call for the highest degree of cognitive processing, and constitute the most challenging "involvement load" to the learners. Their hypothesis was corroborated by the results that the composition group scored significantly higher than the gap-fill and the reading group for both short-term and long-term retention of the new vocabulary. It should be noted that Hulstijn and Laufer also concluded that "involvement load" is independent of receptive or productive language processing.

Following Hulstijn and Laufer's Involvement Load Hypothesis, Kim (2011) investigated why certain types of tasks are more effective in promoting L2 vocabulary acquisition. By comparing two types of tasks (writing the original sentences vs. writing an essay with the target vocabulary), she concluded that vocabulary-focused composition tasks induced a high involvement load, and led to new word learning and retention. However, Kim didn't try to establish any correlational relationships between L2 vocabulary acquisition and language development in complexity, accuracy and fluency. In other words, writing activities in Kim's study were merely employed to examine the occurrence and retention of the target words, and a discussion of the assessment of the writing performance was not included.

To date, studies involving "the involvement load hypothesis" have been mainly concerned with new or the incidental learning of new vocabulary. No studies have yet extended the analysis of the "depth of processing" (Hulstijn & Laufer, 2001, p. 5) into other types of vocabulary acquisition, for instance, formulaic language acquisition. It is not hard to ponder the possibility that directing learners' focus on formulaic language will actually influence the ways attention is allocated (as it does for single words in Kim's study), and will contribute to learners' retention and production of such linguistic forms in their immediate and possibly delayed written products.

Two recent studies (Bitchener & Ferris, 2012; Sheen, 2011) also examined vocabulary acquisition through writing instruction as a pedagogical approach. However, even if vocabulary was the focus in the research, the role of pre-writing planning was not fully addressed given that planning does not seem necessary for every writing study. It is yet to be explored whether planning with a focus on linguistic forms can be used as a pedagogical approach to promote the acquisition of formulaic language.

Rott's (2009) study is one of the earliest studies that investigated planning effects, through an awareness-raising task, on the use of formulaic constructions by learners of L2 German for a written task. Her conclusion is that awareness-raising activities (such as pre-task brainstorming) prior to the writing influence learners' usage of target formulaic constructions in their writing positively and thereby provide an opportunity for learning. Her results are of substantial importance because she addressed the language typology influence (e.g., formulaic constructions are frequently not adjacent in German unlike those in English) on formulaic language acquisition, while the majority of studies targeted English as the L2.

The current study focused on formulaic language in Chinese as a second language (L2 Chinese). Chinese has been regarded as a critical and strategic foreign language in the United States for various reasons, and the need for Chinese learners with balanced Chinese language skills, including listening, reading, speaking and writing, is growing. However, writing has received the least amount of attention as compared to the other skills in L2 Chinese classes. Furthermore, writing in L2 Chinese is commonly practiced as an individual act. Learning to write in Chinese and writing practice are usually left to students themselves because teachers worry that teaching writing in class consumes a lot of class time (Liao, 2010). Although planning in writing and a lexical approach to writing instruction are fairly new concepts in L2 Chinese classes, a few researchers (e.g., Liao, 2010; Yuan, 2010) have started researching the role of planning and effective writing instruction for L2 Chinese. Given a multitude of studies on the definition and categorization of Chinese formulaic language (Kim, 2012; Ma; 2010; Song; 2009; Wang, 2013; Zhan, 2012; Zhou, 2009), there is an urgent need to focus on the teaching and learning of L2 Chinese formulaic language.

In the dissertation, I propose and use the term "a lexical approach to writing instruction" with the attempt to connect three frameworks in SLA research: FS use and uptake, task planning, and writing complexity, accuracy and fluency (CAF) measures. Overlapping of the theoretical frameworks generates some intriguing questions: What are the roles of planning condition and guided planning in formulaic language use and uptake? Do the assumptions about task planning and vocabulary use (for single words in previous studies) apply to formulaic sequences (FSs) as well? Will directing learners' attention to the target FSs in the planning stage increase their (accurate) usage of such forms in their

written outputs? If so, what factors contribute to learners' allocation of attentional resource during writing? All these questions will be reviewed in the literature and corresponding research questions and hypotheses will be stated.

1.2 Rationale of the study

The primary focus of the study is on planning, especially the role of guided planning in written production. Much has been discussed about types and effectiveness of planning in second language acquisition research (Bygate, 1999; Ellis, 2003, 2009; Robinson, 2003, 2005; Schoonen et al., 2003; Skehan & Foster, 2001; Tavakoli & Foster, 2008). As a matter of fact, planning is "one of the pedagogic techniques most widely studied in the task-based teaching literature to date" (Mochizuki & Ortega, 2008, p. 14). The argument is that if learners are given the opportunity to plan, they are more likely to improve in terms of language complexity, accuracy and fluency in task-based performance. With regard to contextual variables (e.g., task instruction and task conditions), in a speaking task Sangarun (2001, 2005) manipulated three types of planning conditions and a group that planned with minimal guidance, and concluded that all three planning conditions (but not the minimal guidance planning) had positive effects on the quality of speech, including both fluency and accuracy. She hypothesized that planning on meaning and form can be simultaneous; however, she warned that learners must have already successfully established form–meaning connections of the target linguistic items (grammar or lexis) before the parallel processing mechanism kicks in.

At the same time, results from previous studies on speaking and writing tasks are mixed as to which aspect of language performance, complexity, accuracy or fluency, will

be influenced the most by task variables. For instance, divergences between Robinson's Cognition Hypothesis (Robinson, 2001, 2003, 2005, 2007) and the Limited Attentional Capacity Model (Skehan & Foster, 1999, 2001) focus on the trade-offs among CAF measures, especially between complexity and accuracy measures. Skehan and Foster argued that learners are not able to attend to both form and meaning, as captured by CAF measures, due to their limited cognitive resources, resulting in trade-offs among CAF measures. In contrast, Robinson's hypothesis claims that more cognitively demanding tasks promote linguistic complexity and accuracy in parallel, but not necessarily fluency.

Recent studies reporting the effects of planning sub processes (Johnson et al., 2012; Ong & Zhang, 2010, 2013) and what really happens during the planning process using think-aloud protocols and retrospective interviews (Mochizuku & Ortega, 2008; Kim, 2013, Park, 2010) allow researchers to scrutinize the contextual and individual variables involved in the planning stage and put forward a more focused research agenda on studying the planning effects. In addition, the confusing definitions and distinctions between task complexity and task conditions (e.g., task instruction) were meticulously discussed in Ellis (2009), Robinson (2011) and Skehan (2009). One of the key arguments in Ellis, Robinson and Skehan's articles is that variations in CAF measures may be attributed to planning conditions and instructions rather than task complexity, per se.

What's more, evidence for the facilitative role of planning has been mainly found in oral language production. Given the paucity of research on the role of planning in written language production, the goal is to find out whether task conditions (e.g., task instruction) contribute to learners' variations in written language production by

manipulating the planning conditions which direct learners' attention to linguistic forms, content, or both during the pretask planning stage.

Finally, few studies of task planning (either pretask or online planning) were set in a regular classroom setting. The current study tested the possibilities of involving pre-writing planning sessions as a routine classroom procedure. Pedagogical significance is prioritized to explore how planning can be manipulated as instruction intervention for formulaic language use and uptake, and written language development. It is hoped that such an endeavor towards research and pedagogy would enlighten future studies of a similar nature.

1.3 Research questions and hypotheses

The ultimate assumption underlying the dissertation study is that directing learners' attention to target L2 Chinese formulaic sequences (FSs) during that pretask planning will prompt students to use such sequences correctly, and consequently affect their written performance in terms of complexity, accuracy and fluency (CAF) measures. To test this hypothesis, four research questions were proposed for the dissertation study:

Research Question 1: What are the effects of manipulating planning conditions, namely, language focused planning (LFP), content focused planning (CFP) and content and language focused planning (CLFP), on L2 Chinese learners' written products in terms of the complexity, accuracy and fluency (CAF) measures?

Hypothesis 1: Participants from CLFP group will not perform similarly, in terms of both accuracy and complexity measures, to those in the CFP group and the LFP groups. This hypothesis derives from Foster and Skehan's Limited Attentional Capacity model

(1999) which stated that content focused planning will lead to greater complexity than planning directed to language, and only language focused planning will lead to higher accuracy. So there could be trade-offs between accuracy and complexity measures as they are both primarily related to L2 knowledge representation.

In addition, one objective of this study was to examine if L2 writers' use of FSs in writing would change according to planning condition. It is widely acknowledged that genre and topic familiarity have an effect on complexity, accuracy and fluency in written tasks (e.g., Sangarun, 2001; Yang, 2014). To reduce the potential effect of having participants write on the same topic multiple times, a second topic was selected that was considered comparable in nature and equally familiar to for the writers, so it is hypothesized that similar performance would be over the CAF measures in both writing sessions.

Research Question 2: What are the effects of manipulating planning conditions on L2 Chinese learners' correct use of the target FSs in writing? Will the participants in the LFP group or the CLFP group (with a focus on both content and language) use more target FSs than the participants in the other two groups?

Hypothesis 2: Participants from both the LFP and CLFP groups will exceed the CFP group in terms of the number and accurate usage of the target FSs. Skehan's model agrees that heightened attention to linguistic forms during the planning stage will lead to increased accuracy of the target structures. In addition, certain types of FSs (multiword structures, but not necessarily phrasal collocations and sentence frames) will be used more accurately, suggesting the influence of the structure of formulaic language such as the degree of "fixedness." In addition, the "all-or-nothing" principle (Schmitt, 2004, pp.

6-14) seems to be able to explain the use of Chinese FSs in a production task, which predicts that FSs will be used either as "holistic units" (p. 4) or not being used at all.

Research Question 3: Are there trade-offs between complexity and accuracy measures? What are the relationships between the CAF measures and learners' usage of the target FS under each of the planning conditions?

Hypothesis 3: As stated in Hypothesis 1, there could be trade-offs between accuracy and complexity measures as they are both primarily related to L2 knowledge representation. Furthermore, participants' retrieval and production of the target FSs is assumed to be related to the fluency (defined as number of words per essay in the study), but not necessarily to the accuracy and complexity measures. Since fluency is more related to learners' control over and automaticity of their linguistic L2 knowledge (Skehan & Foster, 1999), it is hypothesized that retrieval of the formulaic language in the L2 mental lexicon will help learners access their exemplar-based system faster, and possibly lead to an increase in fluency, but not necessarily in complexity or accuracy.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

This chapter reviews both the theoretical claims and empirical studies concerning the three frameworks in the study: task planning, formulaic language use and writing complexity, accuracy and fluency (CAF) measures. The review concerns the nature and characteristics of planning and the role of planning in oral and written production. Regarding the interactions among CAF measures, Skehan's Limited Attentional Capacity Model (1990, 2001, 2009) and possible trade-offs between CAF measures are introduced and compared with regard to differences in planning conditions.

Next, definition and operationalization of formulaic sequences (FSs) are elaborated, followed by a synthesis of empirical studies on the role of formulaic language in L2 learning and teaching, and specifically on the role of planning in formulaic language use.

The last part of the literature review is devoted to introducing Chinese formulaic language and syntactic features in Chinese, including topic-comment constructions and topic-chains. Text analysis in the current study was conducted based on terminal topic-comment unit (TTCU) L2 Chinese. The rationale for adopting such a measure for text analysis is also discussed.

2.2 Planning

The role of planning in language use is not in doubt. Mochizuki and Ortega (2008, p. 12) noted that "it is one of the pedagogic techniques most widely studied in the task-

based teaching literature to date." In fact, there are two trends of studies involved in the Task-based language teaching (TBLT) framework. A majority of TBLT studies (e.g., Ellis, 2003, 2005; Ortega, 1999, 2007; Robinson, 2003, 2005; Skehan, 1998) focuses on task complexity in affecting language production. What is equally important is the role of pre-task planning in TBLT framework. The underlying hypothesis is whether and how planning will help students improve their language use in terms of complexity, accuracy, and fluency. While tasks serve as basic units for syllabus design, classroom methodology and language assessment in TBLT, planning is also an indispensable component in the TBLT framework (e.g. Bygate, 1999; Ellis, 2003; Robinson, 2003, 2005; Schmidt, 1995; Skehan & Foster, 2001; Tavakoli & Foster, 2008).

Planning has been known as explicitly focusing learners' attention on either content or on their choice of language (Ellis, 2005), or both (Ortega, 2007). For instance, Nunn (2006) expresses his preference for planning on linguistic forms as a beneficial procedure in TBLT. Littlewood (2004) addressed the dichotomous nature of "task" and categorized tasks as focus on form or focus on meaning. Ortega (2007) argues that language learning is fostered when learners concurrently attend to meaning and form.

Recent studies on planning have generated a comprehensive view of the various aspects of planning and their effects on learners' L2 output, such as the availability of planning time (Wigglesworth, 1997; Wigglesworth & Edler, 2010), the length of planning time (Crookes, 1989; Mehnert, 1998) , pre-task planning versus online planning (Ellis & Yuan, 2004; Yuan & Ellis, 2003), and written versus oral planning (Kawauchi, 2005). For instance, Kawauchi (2005) explored the effects of three planning modalities on oral production: writing, rehearsal and reading. Although there were no statistical

differences between the three types of planning on task performance, she found that the high proficiency EFL learners benefited from planning the most in terms of fluency, while the low proficiency learners had an increase in accuracy.

One line of research (e.g., Foster & Skehan, 1996, 1999; Kim, 2013; Mochizuki & Ortega, 2008; Park, 2010) discusses planning as instruction intervention and its effectiveness in SLA. Foster and Skehan (1996) compared the inclusion or exclusion of guidance to participants on their use of planning time, while the second one investigated whether the change of focus (form or content) in pre-writing planning had an effect on writing performance. In Foster and Skehan's (1999) study, they directly addressed what inspired the current study the most: how the shift of focus on language forms or content in planning might affect language production. As Yuan and Ellis (2003, p. 3) pointed out, "the studies to date provide very little information about what participants did while they were performing the task." Similarly, very little was known about the role of planning.

It is believed that the role of planning in task-based performance is not only of theoretical interest, but of pedagogical significance because planning allows L2 teachers to design favorable conditions to elicit language output (Foster & Skehan, 1996; Skehan & Foster, 1997, 1999; Ortega, 1999). In other words, planning itself can be used as an instructional intervention approach.

2.1.1 Focus on form vs. content in planning

Two themes emerge in the literature regarding task planning: 1) guided and unguided planning; and 2) the "focus" of planning. With regard to the first distinction between guided and unguided planning, Foster and Skehan's (1996) study was the first to

compare detailed (with instructional guidance and teacher's involvement) and undetailed pre-task planning. They found that detailed planning significantly promoted speech complexity and fluency (but not speech accuracy) as compared with the undetailed pre-task planning group. In a subsequent study, Foster and Skehan (1999) compared teacher-led, solitary and group-based planning on oral task performance. They concluded that teacher-led planning helped learners achieve greater accuracy, while the solitary planning condition led to an increase in complexity and fluency. Students' performance during group-based planning was not as superior to that of the control group which was not given any opportunity to plan. In Mochizuki and Ortega's (2008) study, participants were asked to focus on the use of a single grammatical structure - English relative clauses – in an oral story-telling task. The researchers tracked learners' attention allocation during the planning processes and found that the guided planning group outperformed the unguided planning group and no planning group in accurate usage of relative clauses, but not on the complexity and fluency measures.

Another theme involved in the investigation of planning types is the "focus" of planning: focus on language or focus on content. Crookes (1989) guided his participants to plan both the content and the language for their spoken performance. It was found that participants gained significantly higher complexity (but not accuracy) under the content and language focused pre-task planning than the minimal pre-task planning condition. Similarly, participants in Wendel's (1997) study were asked to plan the content, vocabulary and discourse structures of their oral narratives. However, Wendel concluded that the content and language focused planning contributed to the development in oral

fluency and complexity (but not accuracy and lexical variety) as compared to the control group.

By examining the availability and length of planning time on a set of oral tasks, Wigglesworth (1997) reported that only high proficiency ESL participants benefited from pre-task planning. She hypothesized that the higher proficiency participants were able to plan both content and language when given planning time, while the lower proficiency participants were likely to plan only content. Mehnert (1998) told her participants to plan what they would say and how they would say it in an oral task. She found that participants improved on all aspects of language performance, including speech accuracy, fluency and lexical range, under the content and language focused planning condition compared to the minimal planning condition.

To explore the favorable planning conditions, Foster and Skehan (1999) further studies the effects of foci for planning (language vs. content) on task performance. However, the researchers found little difference in terms of complexity and accuracy between the language-focused and content-focused planning conditions. In other words, both planning conditions failed to direct students' attention "(either) to the language itself, or the ideas to be expressed" (p. 236); however, what matters was the source of planning: whether it was teacher-fronted, group-based or solitary.

Two notable subsequent studies on the effects of different foci of planning in oral production were carried out by Sangarun (2001) and Park (2010). Sangarun modified the planning conditions into language focused planning, content focused planning, and language and content focused planning, and put a major role for the teacher-generated, planned focus on form in the design of her experiments. She reported that all three

planning conditions had positive effects on the quality of speech, including both fluency and accuracy. She hypothesized that planning on content and planning on form can be parallel processing mechanisms.

By examining lexical or grammatical language-related episodes (LREs), Park (2010) investigated whether pretask instructions and planning promote focus on form during task-based interaction. ESL learners participated in oral picture narrative tasks in dyads under one of four conditions: specific instructions with pretask planning, specific instructions without planning, general instructions with planning, and general instructions without planning. Park concluded that the learners prioritized vocabulary use regardless of pretask instructions and planning. In addition, Park pointed it out that it was the task instruction, but not planning per se, that has a role in manipulating attention to form.

In summary, the above discussed studies are inconsistent as to the role of shifting the focus of planning (content focused vs. language focused) in speech complexity, accuracy, and fluency. Nevertheless, this range of studies suggests that when planning is being used to manipulate the ways attention is distributed, and it provides a number of pedagogic choices. The potential of various planning conditions as instructional approaches as summarized in Ellis (2005, 2009) and Mochizuki and Ortega (2008) suggests that an appropriate planning condition would create favorable opportunities for focusing on meaning and form for a pedagogical balance.

2.2.2 The Limited Attentional Capacity Model vs. the Cognition Hypothesis

As discussed previously, planning types and foci differ in how attentional resources are used during task completion. A number of empirical studies (e.g., Ishiwaka,

2007; Kuiken & Vedder, 2007, 2008; Ong & Zhang, 2010, 2013; Skehan & Foster, 2001; Tavakoli & Foster, 2008) have tested the role of attentional resources with regard to different task demands. However, findings are contradictory.

One of the arguments is that if learners are asked to plan, it is likely that there is a joint increase in terms of linguistic complexity and accuracy in task-based performance because focused attention on linguistic forms will likely contribute to the accurate usage of more complex structures and lexical complexity at the same time. The hypothesis has been put forward by Robinson's Cognition Hypothesis (2001, 2003, 2005, 2007). Robinson defines task complexity in a two dimensional category: 1) the resource-directing dimensions, including few/many elements, here-and-now /there-and-then, with/without reasoning demands, and 2) the resource-dispersing dimensions, including with/without planning, single/dual task, with/without prior knowledge. With regard to resource-dispersing dimensions, he argues that planning is such an important component in his task complexity framework that if we take planning time and prior knowledge away from the task, L2 learners' attentional resources will simply be dispersed and consequently, learners are deprived of the opportunities to access their already established repertoire of language.

As a competing theoretical framework on task complexity in task-based SLA research, Skehan's Limited Attentional Capacity Model (Skehan & Foster, 1999, 2001) predicts that because learners' attentional resources are limited, concurrent attention to task content and language forms is difficult to achieve. In other words, when the task demands a considerable amount of attention to its content, there will be only a little attention left for language forms. The hypothesized mechanism is that there are "tensions

between a concern to be fluent, a concern to be conservatively accurate, and a concern to take risks and use more complex language [which] need to be balanced" (Foster & Skehan, 1999, p. 237). When it comes to CAF measures, there should be observed trade-offs between, first of all, meaning (fluency) and form (complexity or accuracy); and then complexity and accuracy are likely to compete with each other. Skehan (1996, 1998) addresses three aspects in determining task complexity: 1) code complexity that includes linguistic and lexical complexity, 2) cognitive complexity that consists of cognitive processing and cognitive familiarity, and 3) communicative stress including time pressure (regarding on-line and off-line planning), modality, and control. In a similar way, Skehan argues that task complexity can be manipulated to direct learners' attention to the complexity, accuracy and fluency of their language output in predictable task conditions. For example, increasing task complexity is likely to draw learners' attentional resources away from linguistic forms, resulting in a decrease in accuracy. In Skehan's model, planning is a separate variable, unlike task complexity, under task conditions (under which tasks are to be done).

Despite the differences in paradigm and predictions between Robinson's Cognition Hypothesis and Skehan's Limited Attentional Capacity Model, it should be noted that both models converge on one point that increasing task complexity along the resource-dispersing dimension (e.g., by taking away planning time, absence of prior knowledge, and increasing the number of concurrent tasks) will negatively affect all aspects - complexity, accuracy and fluency - of language production. However, Robinson does not predict any type of trade-offs, especially between complexity and accuracy, as hypothesized by Foster and Skehan (1996).

2.2.3 Trade-offs between complexity, accuracy and fluency measures

As has been discussed, both Robinson's and Skehan's frameworks aim to explain how task characteristics can affect learning processes and outcomes, and a majority of SLA studies have employed complexity, accuracy, and fluency (CAF) measures to investigate task-based L2 language production. The CAF indices are believed by many researchers (e.g., Ellis 2003, 2009; Ellis & Yuan, 2004; House & Kuiken, 2009; Housen, Kuiken & Vedder, 2012; Palotti, 2009; Skehan, 2009; Skehan & Foster, 1999; Yuan & Ellis, 2003) to be useful for capturing the multiple dimensions of L2 performance. As such, CAF have been used for the oral and written assessment of language learners as well as in interlanguage development studies. Skehan and Foster (1999) defined fluency as "the ability to use language in real time, probably drawing on more lexicalized systems" (p. 96). Accuracy refers to "the ability to avoid errors in performance" (p. 96) and complexity is "the capacity to use more advanced language. This may also involve a greater willingness to take risks, and it is also taken to correlate with a greater likelihood of restructuring" (p. 96-97). Thus, complexity and accuracy are primarily related to L2 knowledge representation, while fluency is more related to learners' control over their linguistic L2 knowledge in terms of the speed and ease with which they retrieve L2 information. A comprehensive review of the development measurement indices for written discourse can be found in Wolfe-Quintero, Inagaki and Kim (1998).

Despite the controversies over the nature and range of complexity, accuracy, and fluency measures, the key to distinguishing Robinson's and Skehan's models lies in whether language complexity, accuracy, and fluency perform simultaneously (as predicted by Robinson), or they vary at the expense of one another (as predicted by

Skehan). On the one hand, trade-offs among complexity, accuracy and fluency, as supported by Skehan's Limited Attentional Capacity Model, were frequently reported in studies such as Skehan and Foster (1997) where L2 learners' oral productions was compared using a narrative task versus a decision-making task. It was found that when given planning time, learners paused less and produced more accurate language in the narrative tasks but not the decision-making task. However, the decision-making task led to an increase in complexity. The authors argued that task characteristics play an important role in "channeling the effect of planning towards accuracy or complexity" (p. 48). Wendel (1997) also found planning led to an increase in fluency and complexity, but not accuracy, in terms of learners' performance in an oral repetition task. He claimed that whether learners attend to aspects of complexity, fluency or accuracy depends on the type of planning, that is, a distinction between what he called strategic planning (pre-task planning) or on-line planning (planning during task performance). By varying the length of planning time given before the writing task, Mehnert (1998) found that fluency improved with each increase in planning time. However, when given more time to plan, participants allocated their attentional capacity to more complex language use, without further improvement in accuracy.

On the other hand, the interaction between syntactic development or complexity and overall accuracy was confirmed in studies such as Bardovi-Harlig and Bofman (1989). Bardovi-Harlig and Bofman examined syntactic and morphological accuracy by advanced English as a Foreign Language (EFL) learners, using the average number of clauses per T-unit (C/T) as the syntactic complexity measure. The idea of including some account for accuracy in the analysis of syntactic complexity may seem contradictory to

the CAF taxonomy (see Wolfe-Quintero, Inagaki, & Kim, 1998 for more discussion); however, Polio (1997) summarized previous studies on syntactic complexity in L2 English and concluded that complexity measures that account for L2 accuracy are the most reliable measurement indices, such as the ratio between the number of error-free T-units and the total number of T-units.

As has been discussed in 2.2.2, both Robinson's and Skehan's models agree that increasing task complexity along the resource-dispersing dimension (e.g., by taking away planning time) will negatively affect all aspects - complexity, accuracy and fluency - of language production. However, complexity or fluency measures are mostly likely to vary in the same direction as accuracy measures according to Robinson, but in opposite directions (for example, increase on complexity but not accuracy or increase on fluency but not accuracy, and vice versa) as predicted by Skehan's model. Yuan and Ellis' (2003) study partially confirmed the unidirectional variation between fluency and accuracy based on Skehan's hypotheses. They compared the effects of pre-task planning, on-line planning and no planning on EFL learners' monologic oral production, and argued for the facilitative role of pre-task planning (defined as unlimited time writing) over on-line planning (defined as timed writing). What their results suggest is pre-task planning promotes fluency and lexical variety, but not accuracy, whereas on-line planning contributes to increased accuracy and syntactic complexity with the potential to inhibit fluency. Yuan and Ellis concluded that learners' focus on the grammatical aspects during online planning would disperse their attention away from content (in other words, what they want to say). Table 1.1 summarizes the CAF measures regarding planning and task complexity in previous studies. Trade-off effects are reported if there were any.

Table 2.1 *CAF measures employed in studies on planning and task complexity*

Study	Written or oral task	Complexity	Fluency	Accuracy	Results
Foster & Skehan (1999)	Oral	number of clauses per C-units	number of reformulations, false starts, repetitions, replacements and pauses; total time of silence	percentage of error-free clauses in overall performance	little difference between the language focused and content-focused groups
Sangarun (2001)	Oral	sentence nodes per T-unit; and clauses per T-unit	syllables per minute; pruned syllables per minute; and percentage of total pausing time	percentage of error-free clauses; and errors per 100 words	An overall effect on fluency, complexity, and accuracy; no differences among the three types (content focused, language focused and content- and language-focused) of planning
Kawauchi (2005)	Oral	number of clauses per T-unit; number of words per T unit; number of subordinate clauses; and number of word types	amount of speech (number of words); percentage of repeated words	use of past forms for copula, regular and irregular verbs	the high proficiency EFL learners benefited the most in terms of fluency, while the low proficiency learners had an increase in accuracy
Mochizuki & Ortega (2008)	Oral	mean length of T-unit; mean number of clauses per T-unit; and number of relative clauses per T-unit	mean number of words per minute	frequency of use of relative clauses; and degree of accurate use of relative clauses	no advantage for guided planning in terms of fluency and complexity; guided planners produced more accurate relative clauses than the unguided planners

Table 2.1 (cont'd)

Ortega (1999)	Oral	words per utterance, type-token ratio, noun-modifier TLU, article TLU	pruned speech rate in syllables per second	target like use of noun-modifier agreement & use of the Spanish article system	mean number of words per utterance and speech rate were significantly higher in the planned output condition; the difference between the mean type-token ratios was not significant; planning effects on the TLU of the noun-modifier agreement was significant, but not the mean TLU of the article system
Ortega (2003)	Written	mean length of sentence, mean length of T-unit, mean length of clause, mean number of T-unit per sentence, mean number of clauses per T-unit, mean number of dependent clauses per clause	n/a	n/a	between-proficiency differences were found in 4 measures: mean length of sentence, mean length of T-unit, mean length of clause & mean number of clauses per T-unit
Ellis &Yuan (2003)	Oral	the ratio of clauses to T-units; total number of grammatical verb forms; mean segmental type-	number of syllabus per minute; number of meaningful syllabus per minute	error-free clauses; correct verb forms	Pre-task planning enhanced grammatical complexity; online planning influenced accuracy and grammatical complexity;

Table 2.1 (cont'd)

		token ratio			more fluency and lexically varied language by pre-task planning
Ellis &Yuan (2004)	Written	the ratio of clauses to T-units; number of different verb forms; mean segmental type-token ratio	Syllables per minute; number of dysfluencies—	error-free clauses; correct verb forms	pretask planning resulted in greater fluency and greater syntactic variety (especially the variety of verb forms), but with a limited effect on accuracy; whereas on-line planning benefited accuracy the most at the price of fluency and complexity
Ishiwaka (2007) *on task complexity*	Written	S-nodes per T-unit; clauses per T; S-nodes per clause; dependent clauses per clause	n/a	target like use of articles	difficult task promoted complexity and accuracy
Kuiken & Vedder (2008) *on task complexity*	Written	clauses per T-unit; dependent clauses per clause	Fluency measures not accounted; instead, lexical variation was coded	total error per T-unit (with three degrees of errors)	more complex task led learners to make fewer errors and use more high frequency words; no significant differences in syntactic complexity and lexical variation; partially support Robinson's Cognition Hypothesis

Table 2.1 (cont'd)

Johnson et al. (2012)	Written	(1) the measure of textual lexical diversity (2) the ratio of pronouns to noun phrases, (3) the incidence of personal pronouns normed to 1000 words, (4) the mean frequency rating with which the content words in a text appear in the English language according to the COBUILD English language corpus, and (5) the normed frequency (per 100 words) of word types from the fourth and fifth most frequent word families	total number of words and average sentence length	n/a	pre-task planning condition had a small significant effect on writing fluency, and no impact on lexical complexity and grammatical complexity; a threshold level of general L2 proficiency may be necessary for pre-task planning to impact L2 writers' texts.

Notes: T-unit = one main clause with all subordinate clauses attached to it
C-unit= a word, phrase, or sentence that in some way contributed pragmatic or semantic meaning to a conversation
TLU=target-like use
type-token ratio= the ratio of different words to the total number of words in a text

2.2.4 Planning in writing

While most of these studies of planning have focused on oral proficiency, there is a paucity of task-based research on the role of planning in written language production. Williams (2011) points out that compared to speaking, writing requires more attention to the formal features of language. The fact that writing is a more monologic type of task suggests it is more cognitively challenging than dialogic interactive tasks (Tavakoli & Foster, 2008). Results from some recent studies on written production of L2 English, French and Italian within the TBLT framework (e.g. Ellis & Yuan, 2004; Ishiwaka, 2007; Kuiken & Vedder, 2007, 2008; Ong & Zhang, 2010, 2013) are supportive of Robinson's Cognition Hypothesis in general, implying that more complex writing tasks will lead to better task performance in terms of accuracy as well as lexical and syntactic complexity.

In a follow-up to their 2003 study with spoken language, Ellis and Yuan (2004) extended their results on the facilitative role of pre-task planning and online planning (both planning conditions were in the written form) to narrative writing by intermediate level EFL learners. They found that pretask planning resulted in greater fluency and greater syntactic variety (especially the variety of verb forms), but with a limited effect on accuracy, whereas on-line planning benefited accuracy the most at the price of fluency and complexity. The trade-off effects may be due to learners' limited processing capacity, so that a gain in fluency or complexity negatively impacts accuracy and vice versa. By drawing on Kellog's (1996) model of writing, Ellis and Yuan discussed why the two types of planning impact different aspects of L2 writing processes. While pretask

planning promotes formulation, unpressured on-line planning allows for more opportunities for monitoring.

As to the effects of manipulating task complexity on written language production, Ishiwaka (2007) compared simple (here-and –now) and difficult (there-and-then) narrative writing tasks and concluded that the difficult task elicited significantly more complex and accurate language production than did the simple task. Writing tasks were also employed by Kuiken and Vedder (2008) to explore the effects of task complexity. They compared predictions generated based on Skehan's model (a negative effect of increasing task complexity on all dimensions of production) and on Robinson's Cognition Hypothesis (increasing complexity along resource-directing variables will improve accuracy and linguistic complexity simultaneously). Dutch learners of Italian and of French at two different levels of proficiency were instructed to write on an easy task (three requirements) and a complex one (six requirements). Learners' written performance was compared in the areas of accuracy (including a classification of grammar, lexical or orthographic errors), structural complexity, and lexical variation (e.g., lexical frequency profile analysis). The results suggested more complex tasks led learners to make fewer errors and use more high frequency words under complex performance. However no significant differences were found in terms of syntactic complexity and lexical variation, which partially supported Robinson's Cognition Hypothesis.

In fact, the emphasis on the role of memory capacity and attentional resources in the writing process during pre-task planning coincides with the shift in focus on cognitive

processes in L2 writing research. Becker (2006) argues that planning, especially in developing outlines for the writing task, will help relieve the burden on working memory for competent writers, while for novice writers, planning skills are to be prioritized as task elements. Regarding the role of planning in writing, De Larios, Marin, and Murphy (2001) asserted that "writers will differ in the way they adapt their time allocation to formulation processes as a result of task conditions " (p. 503).

One underlying assumption for the role of planning in language production is that learners' attention could be directed to their prior knowledge, be it content knowledge or linguistic forms, as required by the task. Nevertheless, how planning aids language learners in specific ways is a matter of controversy (Sangarun, 2005). It seems that both content knowledge and linguistic forms are inclusively grouped under "prior knowledge" along the resource-dispersing dimensions in Robinson's task complexity model, which prompts the current study to go further to explore the differentiated role of planning on content versus planning on linguistic forms in written language output.

2.3 Formulaic language in SLA

There is a growing awareness that a large part of the discourse that we produce is composed of formulaic language (Wray, 2002, 2008). Hopper (1998) stated that discourse abounds in all sorts of repetitions that have nothing to do with grammar: idioms, proverbs, clichés, formulas, favored clause types, and so on. Their boundaries may or may not coincide with the constituent boundaries of traditional grammatical

descriptions.

In adult L2 acquisition, formulaicity is claimed to be closely related to the development of fluency (Oppenheim, 2000; Wray, 2002). Formulaic sequences are also found to be one of the strongest discriminators between lower and higher levels of L2 English proficiency (Laufer & Waldman, 2011, regarding verb-noun collocation in L2 writing; Verspoor, Schmid, & Xu, 2012, regarding lexical chunks in L2 writing). As such, it is hypothesized that ultimate fluency in L2 is achieved by "clause-chaining" or stringing together a sequence of relatively independent clauses (Pawley & Syder, 1983, pp. 203-204) due to the limited memory space of human beings. In other words, memorized chunks are capable of compensating for limitations in the human brain (Wray 2002; Wray 2008). A number of psycholinguistic approaches have been adopted to explore the representation and processing of FSs in the mental lexicon, including, for instance, an elicited imitation test (Schmitt et al., 2004), a priming test (Sonbul & Schmitt, 2013), an online grammaticality judgment task (Jiang & Nekrasova, 2007; Yomasita & Jiang, 2010), a line-by-line self-paced reading task (Conklin & Schmitt, 2007), eye-tracking (Underwood, Schmitt & Galpin, 2004; Siyanova, Conklin & van Heuven, 2011; Siyanova, Conklin & Schmitt, 2011) and ERP measures (J. McLaughlin, Osterhout & Kim, 2004).

2.3.1 Definition and operationalization of formulaic sequences

The definition and categorization of FSs is a critical factor in data analysis, and it

may explain why certain types of FSs are produced more frequently and accurately by L2 learners of English (Schmidt & Carter, 2004). A formulaic sequence can range from completely free syntactic constructions (e.g., *NP be-TENSE sorry to keep-TENSE NP waiting*) to completely fixed or frozen expressions such as *by and large* (Howarth, 1998). Thus a mixture of terms has been attached to the definition of formulaic language, including “lexical phrases”, ”multi-word units”, “formulas”, “prefabricated chunks”, “ready-made utterances”, and so forth (e.g. Nattinger & DeCarrico, 1992). Wray (2002) elaborated her definition of a “formulaic sequence” as:

> a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory, rather than being subject to generation or analysis by the language grammar. (p. 9)

Becker (1975) was the first to propose a six-category-taxonomy of formulaic language in English, which includes polywords (e.g., *for good*), phrasal constraints (e.g., *by sheer coincidence*), meta-messages (e.g., *that’s all*), sentence builders (e.g., *a long song and dance about something*), situational utterances (e.g., *how can I ever repay you?*), and verbatim texts (e.g. *better late than never*) (adapted from Becker, 1975, p. 6). However, he failed to capture the degree of fixedness of the formulaic language, from fixed to partially-fixed to open structures, in a continuum.

Bolinger (1976) proposed the idea of categorizing formulaic language in English based on the extent of its fixedness, from non-idiomatic to idiomatic. Being non-

idiomatic means substitutions of the components are allowed within the construction, while idiomatic constructions permit no substitutions. Under this approach, he summarized four types of formulaic language: free combination (e.g., *under the table*), restricted collocation (e.g., *under attack*), figurative idiom (e.g., *under the microscope*), and pure idiom (e.g., *under the weather*) (adapted from Bolinger, 1976, p. 27-28). Bolinger's categorization mainly involved lexical phrases. He didn't particularly address formulaic constructions at the sentence level.

Howarth's (1998) continuum model is the most cited one for research on formulaic language in English. In his continuum, there are functional expressions (sequences with a discourse role such as openers, proverbs, slogans and so on); composite units (which retain a syntactic function); lexical collocations (consisting of two open class items, such as *ulterior motive*); grammatical collocations (consisting of one open and one closed class item, such as *in advance*) (adapted from Howarth, 1998, pp. 27-40). Howarth addressed the lexis-syntax interface and suggested that all non-institutionalized phrases, even if they might be highly frequent in discourse, are not necessarily retrieved holistically.

Formulaic sequences were referred to as "lexical phrases" by Nattinger and DeCarrico (1992). In their comprehensive taxonomy and description of formulaic language in English, there are four large classes of lexical phrases: Polywords are phrases that operate as single words, allowing no variability (e.g., *for the most part*, *in a nut-shell*, *by the way*). Institutionalized expressions are continuous, sentence-length expressions

which are mostly invariable (e.g., *nice meeting you*, *long time no see*). Phrasal constraints refer to lexical and phrase categories that are mostly continuous with a certain degree of variation (e.g., *a very long time ago*, *as I was saying*). Finally, sentence builders include lexical phrases that serve to construct sentences with fillable slots (e.g., *I think that it's a good idea that…*, *not only…but also…*). What is noteworthy about Nattinger and DeCarrico's model is that the authors also provided a functional analysis of the four types of lexical phrases they proposed. The functions they discussed included: social interaction markers (e.g., *I'm sorry but…*), topic markers (e.g., *my name is…*), discourse devices (e.g., *to make a long story short*) and fluency devices (e.g., *you know*, *so to speak*). The clear distinction between form- and function-based categories contributes to the understanding of the fuzzy nature of formulaic language and allows for further studies to come up with more operational categorizations for various research purposes.

Schmitt and Carter (2004) readdressed the degree of fixedness of frame structures (with fillable slots) in L2 English learners' processing of formulaic language. Using the sentence stem "*thinks nothing of …"* as an example, they argued that this partially fixed structure with a fillable slot allows flexibility of expression in a wide variety of situations. Nevertheless, they cautioned that there are semantic constraints for the slot in this structure that control which word or words can be used. For instance, "*thinks nothing of …*" presumes ideas that are unusual or unexpected. As such, a sentence that reads "She thinks nothing of sleeping eight hours per night" sounds strange because "eight hours per night" does not meet the semantic requirements for the slot to be something unusual or

unexpected. In short, it is generally agreed upon that "fixedness" has been applied successfully as one of the criteria in the categorization of formulaic language in English. In the current study, Nattinger and DeCarrico's (1992) "continuum of fixedness" model serves as the theoretical basis for categorization.

2.3.2 Planning and FS use

A few experimental studies have explored the role of planning in the acquisition of formulaic language by L2 learners (Boers, Eyckmans, Kappel, Stengers & Demechleer, 2006; Bolander, 1989; Foster, 2001; Rott, 2009). Bolander (1989) looked at memorized chunks in free speech and spoken response to a guided task by learners of Swedish. He hypothesized that learners' errors in memorized chunks are the result of their fossilized incorrect forms from earlier usage. Foster (2001) examined the oral production of native and nonnative speakers' classroom interaction with additional data from her two previous studies (Foster & Skehan, 1996; Foster, 2000). For the planning group, students were given a written list of suggestions on vocabulary and grammatical structures they might find useful later in the oral task. She found that non-native speakers at a lower proficiency level of English were limited in producing lexicalized phrases than native speakers, indicating that nonnative speakers were using "a rule-based approach to language production" (p. 90), regardless of the planning conditions.

Boers and his colleagues (2006) measured the effectiveness of pre-task phrase-noticing activities on learners' usage of the target formulaic phrases on oral production.

The authors stated that “no ‘hard’ empirical evidence of the effectiveness of chunk-noticing has been published yet” (p. 248) and their attempt was to draw learners’ attention to formulaic phrases through a pedagogical task. Students in the experimental group were given explicit explanation and examples on word combinations during an “exploration stage in dealing with reading texts or audio/video recordings” (p. 250). Possible formulaic phrases were blind-judged by two native-speaker instructors and only correctly formed chunks were counted for analysis. Their results suggested students benefited from the oral tasks to raise their awareness of the formulaic sequences, and subsequently to recognize the learned sequences in new texts and to reproduce them in conversations.

Rott’s (2009) study is the only one that examined the effects of planning, through an awareness-raising task, on the use of formulaic constructions by learners of L2 German for a written task. She compared students who had the opportunity to brainstorm their ideas to those who brainstormed at least 10 expressions they might want to use later. She found that students who had the opportunity to write down expressions used more grammatically and semantically correct constructions in the task of writing a recipe. She concluded that even though advanced learners experienced obvious difficulties with the target formulaic constructions, awareness-raising activities (such as pre-task brainstorming) prior to the writing influenced learners’ usage of formulaic constructions positively and thereby provided an opportunity for learning. Her study also highlights that in contrast to English, German poses greater challenges to the control of formulaic

constructions since items are frequently not adjacent in German. Her examination of the influence of language typology on formulaic language constitution and acquisition provides important information on the operationalization and categorization of L2 Chinese formulaic language in the current study.

Another interesting point in Rott's study is the measure of multi-word units. In addition to the total number of target constructions used and the total number of correctly produced constructions, she also tallied what Fitzpatrick and Wray (2006) termed as "completeness/accuracy: the number of words produced with the same form and function as in model target utterance ÷ number of words in model target utterance" (p. 46), which is considered as a useful measure for partially produced constructions. The model target utterances were given to the participants during pre-task planning. The reason to count words only with the "same form and function" was to exclude instances where the word happened to be identical to the target one, but was engaged in a different morphosyntactic relationship, such as "*to* as infinitive marker and *to* as a preposition" (p. 46). However, it should be noted that Rott didn't provide further information about the expressions or constructions which learners wrote down during their pre-task brainstorming. In addition, Rott didn't control time as an intervening variable in her study (i.e., participants were asked to write without a time limit), leading to a vague interpretation of the variations concerning the various lengths of texts that the students produced.

Based on the literature, it can be concluded that both task conditions and proficiency contribute to the variances in the retrieval and production of FSs. The

subsequent question to be answered is which instructional design and planning conditions will provide the optimal opportunities for FS use and production.

2.4 Chinese formulaic language and the syntactic features of Chinese

2.4.1 Formulaic language in Chinese

Formulaic language is pervasive in the Chinese language with most multi-word collocations and some grammatical constructions in Chinese exhibiting a certain degree of formulaicity (Su, 2010, 2011). Kim (2012) analyzed and summarized FSs that occurred in the textbook *Boya Chinese* (Li, 2005) for intermediate L2 Chinese learners. She estimated that there was a total of 1,624 FSs in the textbook, with 915 FSs at phrasal levels (accounting for 56.34% of the total FSs). For instance, the phrase 十之八九 *shízhībājiǔ* (which literally means “eight or nine out of ten” and metaphorically, “most probably”) has a semantically comparable counterpart 十个中间的八九个 *shígè zhōngjiān de bājiǔ gè*. Although the latter expression is perfectly understandable to native speakers of Chinese, it lacks the formulaicity or the holistic nature of its formulaic counterpart. It is very unlikely that native speakers of Chinese would go through (in speaking or writing) each individual morpheme in the second phrase 十个中间的八九个 *shígè zhōngjiān de bājiǔ gè*, given the presence of the cognitively more efficient expression 十之八九 *shízhībājiǔ* in their mental lexicon. The amount of cognitive effort that native speakers save by using formulaic language tends to put non-native speakers at a disadvantage when they try to express ideas with the non-formulaic equivalents of the

formulaic expressions. In some cases, non-native speakers are even unaware of the existence of such formulae.

Following Wray (2002) and the theoretical frameworks of formulaic lexical and syntactic structures in Chinese (Kim, 2012; Ma; 2010; Song; 2009; Wang, 2013; Zhan, 2012; Zhou, 2009), the working definition of FS in the current study is operationalized as a set of characteristics: a) FSs are strings of words that belong together according to native speakers' intuition; b) they coexist in a fixed/partially-fixed/open continuum; and c) they have a relatively high corpus frequency.

Furthermore, the categorization of Chinese FSs has been discussed in the literature with reference to the compositionality of meaning, syntactic structure (phrasal vs. sentential), internal structure, grammatical well-formedness, communicative function and metaphoricity (for idioms and proverbs especially). Gao (2008) proposed a framework with five categories along a continuum of fixedness, ranging from partially-fixed lexical phrases to open structures with fillable slots: multi-word collocations (e.g., 仅次于 *jǐncìyú* [only next to]), parentheses (e.g., 看起来 *kànqǐlái* [looks like]), phrasal frames (e.g., 跟…见面 *gēn … jiànmiàn* [to meet with…]), sentence cohesive devises (e.g., 之所以…是因为… *zhīsuǒyǐ … shìyīnwéi …* [the reason for… is because…]), and special sentence structures (e.g., V着V着就…*V zhe V zhe jiù…* [doing something leads to other (often unexpected) consequences]). Although this five category division is rather inclusive, the distinction is vague in terms of their grammatical functions of each type of formulaic language.

Zhan (2012) specifically focused on the frame structures in Chinese and proposed a categorization of FS at the word level, phrase level and sentence level. Examples of four-word frame structures include 爱X不X (e.g., 爱理不理 *ài lǐ bù lǐ* [pay no attention to], here both X and Y are verbs), 没X没Y (e.g., 没大没小 *méi dà méi xiǎo* [not disciplined], here both X and Y are nouns), 不X不Y (e.g., 不三不四 *bù sān bù sì* [not moral], here both X and Y could be adjectives), 大X大Y (e.g., 大鱼大肉 *dà yú dà ròu* [a feast], here both X and Y are nouns), X东Y西 (e.g., 说东道西 *shuō dōng dào xī* [beating around the bush], here both X and Y are verbs) etc. Phrasal level frames refer to formulaic patterns such as 当… 的时候 *dāng ... de shíhòu* [when…]，对…来说 *duì … lái shuō* [meaning...to…]，以…为中心 *yǐ ... wéi zhōngxīn* [centered around…]，对…感兴趣 *duì ... gǎn xīngqù* [be interested in…]，要是…的话 *yàoshì … de huà* [if.. happened]. As to sentence level frame structures, both cohesive devices/connectives (e.g., 不但…而且… *bùdàn ... érqiě ...* [not only… but also…], 虽然…但是… *suīrán ... dànshì* … [although… but…]) and some special sentence patterns (e.g., 把字句 ba-structure) are included. Zhan's categorization seems most feasible in the way that it captures the degree of fixedness at three levels; however, the limitation of the taxonomy is that it is limited to analyzing frame structures in Chinese, to the exclusion of other types of FSs.

J. Zhou (2007) classified formulaic language in Chinese as fixed collocations, idiomatic phrases and fixed structures such as sentential connectives. Adding one more category of "frame structures" based on J. Zhou's (2007) work, Q. Zhou (2009) classified

a total of 860 chunks in her study into four categories: idiomatic collocations, fixed phrases, frame structures and sentence patterns (including sentential connectives).

Following J. Zhou (2007) and Q. Zhou (2009), three types of Chinese FSs were proposed for analysis and comparison in the current study, namely, multi-word collocation, phrasal frames and sentential connectives. Among the three categories, *multi-word collocation* is the most "fixed" type based on Nattinger and DeCarrico's (1992) "continuum of fixedness." A fully fixed FS implies that it doesn't allow random choices for its morphological constituents (Moon, 1997). The second type is *phrasal frames* which refer to partially fixed structures with fillable slots. The third type, *sentential connectives*, includes open structures that are the least "fixed" grammatically. The three types of FSs in the study corresponded with Nattinger and DeCarrico's polywords, phrasal constraints and sentence builders respectively.

It should be noted that one important category of formulaic language, specifically idioms, proverbs, and sayings, was deliberately excluded from the analysis in the current study. The reason is that idioms, proverbs, and sayings are strictly fixed and easily identifiable by their forms. What is more, they manifest obvious processing advantages (in terms of retrieval and production) as single units over other types of FSs.

2.4.2 Topic-comment structures and topic chains in Chinese

Typologically, Chinese and most Indo-European languages differ as to the role of subject and topic. English is a subject-prominent language, and a full sentence in English

is characterized by the presence of a subject and a predicate. A subject in an English sentence is formed by a noun phrase (NP) which plays a prominent role. However, the subject position in Chinese can be a NP, topic phrase (TP), adverbial phrases (AP) and prepositional phrases (PP), an entire clause, or even an empty NP (Li & Thompson, 1976, 1981). Similarly, the predicate in a Chinese sentence may lack a finite verb. Furthermore, a NP itself without any predicate can stand alone as an independent sentence in Chinese. In short, Chinese syntactic patterns can be simply classified into four types from the perspective of subject: sentences with a topic and subject, sentences with a subject, sentences with a topic, and sentences with neither topic nor subject.

Li and Thompson (1981) pointed out that Chinese sentences are built upon semantic-based word order rather than agreement morphology. Consequently, sentence-hood in Chinese is not as transparent as it is in English. The grammatical relationship between a subject and a predicate has been defined as that of topic and comment in Chinese. Norman (1988) argued that topic-comment constructions are more prevalent than subject-predicate formations in Chinese, and are found both in spoken and written discourse. Example (1) shows a typical TP construction in Chinese. In this example, the topic of the sentence is “this sentence” and the comment is “I have to say.”

(1) 这句话我不得不说。

Zhè jù huà wǒ bù dé bù shuō.

This quantifier sentence I have to say.

(*I have to say this.)*

Furthermore, Yip (1995) classified topic structures in Chinese into two categories: derived topics and base-generated topics, and concluded that English has lower acceptability towards moving a non-subject to the topic position.

(2) 北京很多人骑自行车。Yip (2005, p. 22)

Běijīng hěnduō rén qí zìxíngchē.

Beijing very many people ride bicycle.

(In Beijing, many people ride bicycles)

(3) 北京我没去过。Yip (2005, p. 22)

Běijīng wǒ méi qù guò.

Bejing I not go Aspect Marker.

(I have not been to Beijing.)

The topic “Beijing” in (2) does not bear any semantic relationship to the subject “people.” This is an example of the base-generated topic structure. However, in (3) the topic “Beijing” represents the object of the verb “go” and has undergone movement from the object position to the topic. A structure like this is called a derived topic sentence. Both types of topic structures are typical of native Chinese speakers’ language.

Another feature of Chinese syntax involves the concept of “topic chain” (Li & Thompson, 1979, 1981; Shi, 1989, 2000; Tsao, 1979, 1990) which refers to a chain of clauses sharing a single topic. In a topic chain, the topic extends its semantic control over the subsequent clauses or sentences within the chain, which, in contrast, would be run-on sentences in English (Xiao, 2004). The topic is usually mentioned once at the beginning

of a chain in the first clause. Subsequent mentions of the same topic are usually left unspecified (Li, 2004a).

(4) 素食的好处很多，可以影响健康，也可以保护环境。

Sùshí de hǎochù hěnduō, kěyǐ yǐngxiǎng jiànkāng, yě kěyǐ bǎohù huánjìng.

vegetarian -de advantage many, Ø can affect health, Ø also can protect environment.

(There are many advantages of being a vegetarian. It can affect your health. It can also protect the environment.)

In (4), there are two unspecified empty positions (marked as Ø) that are coreferential with the overt subject "Sushi" in the preceding text. The Øs are also referred to as zero anaphora.

Syntactically, the two clauses with a zero anaphor represent what Li and Thompson (1979) called "a massive non-specification of arguments" (p. 317), and which is unacceptable in SP languages such as English. However, topic chains in Chinese account for most of the use of zero NPs, and Chinese makes much less use of anaphoric pronouns (Li, 2004a). Moreover, a Chinese topic chain can consist of more than one controlling topic or coreferential relationship. Based on the number of controlling topics or coreferential relationships, topic chains can be classified as single-link, double-link, or triple-link (Xiao, 2004). Example (5) is a double-link chain with two controlling topics:

(5) 今天早上他在报纸上看到一个广告，

Jīntiān zǎoshang tā zài bàozhǐ shàng kàn dào yīgè guǎnggào,

this morning he at newspaper saw an advertisement,

说学校附近有一个公寓出租，

shuō xuéxiào fùjìn yǒu yīgè gōngyù chūzū,

Ø said school near there-was an apartment on lease,

离学校只有一公里，很方便。(Xiao, 2004, p. 137, from Yao & Liu, 1997)

lí xuéxiào zhǐyǒu yī gōnglǐ, hěn fāngbiàn.

Ø from school only have one kilometer, Ø very convenient.

(He saw an advertisement in the newspaper this morning. It said there was an apartment for rent near the school. The apartment is only one kilometer away from the school which is very convenient.)

In (5) there are two instances of zero anaphora; however, both of them are not coreferential with the overt subject "he." Instead, the controlling topic for the first clause is "advertisement", and "apartment" controls the following clause. In other words, there are two coreferential relationships. Although the coreferential relationships in (5) are not hard to capture by native speakers of Chinese, a sentence like this would be regarded as a run-on sentence in English.

2.4.3 Terminal Topic-Comment Unit & TTCU based text analysis

Even though complexity, accuracy, and fluency (CAF) measures have been widely employed (e.g., Ellis 2003, 2009; Ellis & Yuan, 2004; Skehan & Foster, 1999; Yuan & Ellis, 2003) to capture the multiple dimensions of L2 performance, quantitative

analysis based on CAF measures would not be possible (Foster et al., 2000) without a unit of measurement. The T-unit was one of the first proposed and one of the most widely accredited and used measurement units for accuracy, fluency and syntactic complexity (see Table 1.1). The T-unit is defined as "the shortest units into which a piece of discourse can be cut without leaving any sentence fragments as residue" (Hunt, 1970, p. 188) and "it is one main clause with all subordinate clauses attached to it" (Hunt, 1965, p. 20). According to Hunt, the use of T-unit based measures allows to distinguish complex sentences from compound sentences (sentences exhibiting main clause coordination, Hunt, 1965, 1970), whereas clauses are identified by verb phrases (that is, as phrases dominated by either VP or S).

Research shows that the language of advanced L2 Chinese learners exhibits a wide range of syntactic variation, including TP constructions and topic chains which represent the language of native Chinese speakers (Jin, 1994; Yuan, 1995; Xiao, 2004). Neverthelsee, there is a paucity of studies analyzing L2 Chinese texts with objective measures (Jiang, 2013; Jin, 1994, 2007; Yuan, 2009, 2010). Given the fact that sentence-hood and clause-hood are vague terms in Chinese, there is a greater flexibility as to the nature of the subject and the predicate than that in subject-prominent languages. However, this flexibility in terms of subject and predicate poses a big problem for sentence-, clause-, and T-unit based text analysis of Chinese. So far, there are only a few studies (Jiang, 2013; Jin, 1994, 2007; Yuan, 2009, 2010) on analyzing L2 Chinese texts with development indices such as CAF measures. As Robinson, Cadierno and Shirai (2009, p.

550) noted “specific measures should be more sensitive to conception, task complexity, and its linguistic demands than general measures,” the field of SLA is in urgent need of objective and reliable measure(s) or measurement units for analyzing L2 Chinese written language.

Jin (1994, 2007) extended the T-unit measure to L2 Chinese texts and argued that although the T-unit is a reliable index for L2 SP languages, it may not be appropriate for Chinese since Chinese is a topic-prominent language. She suggests a new index: Terminal Topic-Comment Unit (TTCU) based on the syntactic features of zero anaphor and topic chains in Chinese. She reported that the mean length of TTCU correlated positively with the development of Chinese writing proficiency. In other words, with the increase of learners’ L2 Chinese proficiency, they tend to write longer TTCUs. Example (7) shows what Jin defined as a typical topic chain, counting as one TTCU.

(7) 素食的好处很多，可以影响健康，也可以保护环境。

Sùshí de hǎochù hěnduō, kěyǐ yǐngxiǎng jiànkāng, yě kěyǐ bǎohù huánjìng.

vegetarian. –de advt. many, / Ø can affect health, / Ø also can protect environment.

(There are many advantages of being a vegetarian. It can affect your health. It can also protect the environment.)

Example (7) consists of three T-units, even though the subject is missing in the following T-units. According to the principle of zero anaphor, all three T-units are considered error free. At the same time, the T-units constitute one TTCU in Chinese.

(8) 公司想要消费者认为他们为了表达爱而需要买东西送给爱人。

Gōngsī xiǎng yào xiāofèi zhě rènwéi tāmen wèile biǎodá ài ér xū yāo mǎi dōngxī sòng gěi àirén.

Company wants customers think they for express love / so Ø need buy something to give in-love person

(The company wants the consumers to think that they are buying something for the person he/she loves to express their love.)

Example (8) is a grammatically correct sentence with zero anaphor. The sentence constitutes a TTCU with 26 characters.

(9) 之所以买东西是因为消费者想表达爱，可是，*这样来说的看法并不对。

Zhī suǒyǐ mǎi dōngxī shì yīnwèi xiāofèi zhě xiǎng biǎodá ài, kěshì,*zhèyàng lái shuō de zhèyàng lái shuō de kànfǎ bìng bùduì.

why so buy sth. is because customers want to express love, / but, so to speak de view Ø at all not right

(The reason for buying is that the consumers want to express love. However, this view is just not right.)

Example (9) is another sentence with zero anaphor; however, the second half of the sentence is grammatically incorrect. The phrase “zhèyàng lái shuō (*so to speak*)” can only be used to connect two independent sentences, but not as a modifier to the noun “kànfǎ (*view).*”

(10) 对我来说，*过如何的节日的最重要的是吃美食。

Duì wǒ lái shuō,*guò rúhé de jiérì de zuì zhòngyào de shì chī měishí.

To me talk, live how de festival de most important de is to eat delicious food

(To me, the most important thing to celebrate a festival is to eat delicious food.)

In Example (10), there is a syntactic error with the phrase "guò rúhé de jiérì *(live how de festival).*" The correct order in Chinese should be "rúhé guò jiérì *(how live festival)"* as the topic of the topic-comment structure. The rest of the sentence "*the most important thing is to eat delicious food"* serves as the comment.

In order to calculate the mean length of TTCUs in examples (8), (9) and (10), the number of characters in each TTCU (26, 28 and 19 Chinese characters respectively) are added up and divided by 3. The mean length of TTCUs is 24 characters.

With regard to CAF measures, Yuan (2009, 2010) looked at the effects of task planning on L2 Chinese oral and written data by L1 English learners. Although she did not use any topic-based measurement units such as TTCU, she suggested including TTCU in future studies of L2 Chinese text analysis. Jiang (2013) used a cross-sectional design to investigate measures for L2 Chinese writing development of three groups of native English speakers. Among the three T-unit based measures she explored, the percentage of error-free T-units was found to be the only measure that discriminated between learners at all levels. Comparing L2 learners' data with those of Chinese native speakers, she did not find a statistical difference among the learners using measures of T-unit length and error-free T-unit length.

It is obvious that there are several issues concerning the validity of using T-unit measures in L2 Chinese text analysis, even though Jiang (2013) claimed that the T-unit is

a reliable measure for different stages in L2 Chinese syntactic development. The Chinese language does not depend on coordinate or subordinate conjunction for syntactic complexity, thus it seems almost impossible to find any long T-units, defined as units of 21 or more words by Hunt (1970, 1976), as those found in advanced L2 English data. As learners' proficiency improves, their written language will exhibit features that represent native speakers' language. In the case of L2 Chinese syntactic complexity, advanced learners are expected to produce a great number of topic-comment structures and topic chains, and these topic chains will be comparable to native speakers' language in terms of average length and accuracy. If the T-unit fails to capture these syntactic complexity features by advanced L2 Chinese learners, the validity of using the T-unit based text analysis in analyzing advanced L2 data should be questioned, although it might be powerful in detecting differences in lower or intermediate L2 data. Given the fact that topic-comment structures and topic chains are so prevalent in native Chinese speakers' language, it is hypothesized that TTCU-based measurements, as compared to T-unit based measurement indices, are more robust in capturing the differences in complexity, accuracy and fluency of advanced L2 Chinese learners' written production.

2.5 Critical evaluation of the previous research

In short, major problems from previous studies can be summarized into three aspects: 1) the role of guided planning in facilitating attention to linguistics forms and

content is not quite clear; 2) research results based on CAF measures were inconsistent; and 3) T-unit based measures were not robust in measuring L2 Chinese texts.

First of all, various pre-task activities (e.g., vocabulary and grammar list in Foster, 2001; noticing activity in Boers et al., 2006; and brainstorming in Rott, 2009) have been employed in previous studies in an effort to direct learners' attention to the target linguistic forms at the planning stage. However, there is a lack of consistency in terms of how FSs are scored and coded for analysis. Also, none of the studies provided further details about what the instructions were for the planning activities and how the instructions were given to the students. Furthermore, in order to control the interference of contextual variables, tasks need to be performed within a time limit. However, neither Boers et al. (2006) nor Rott (2009) reported that time was controlled as a variable in their studies. Finally, as suggested by Foster (2001), it should be interesting to look at the range and categories of the formulaic phrases produced in L2 production.

Another issue concerns the inconsistency among CAF measures. Housen, Kuiken and Vedder (2012) argued that "when it comes to the usefulness and validity of complexity, accuracy and fluency as research constructs, this is where the consensus ends and the controversy begins" (p. 300). Michel, Kuiken and Vedder (2012) called for researchers to include task specific measures rather than global CAF measures when discussing statistical significance of the used measures related to the Cognition Hypothesis. For example, Foster and Skehan (1999) reported little difference between the language-focused (modal verbs and conditionals) and the content-focused group in terms

of complexity and fluency. However, Mochizuki and Ortega's (2008) study suggested that guided planning contributed to an increase in terms of accuracy. A closer examination between the accuracy measures in these two studies revealed that Mochizuki and Ortega used a very targeted measure of accuracy (on the frequency of use of relative clauses and the degree of accurate use of relative clauses), whereas Foster and Skehan (1999) only used a general accuracy measure (percentage of error-free clauses in overall performance). It is no wonder that such a general measure of accuracy failed to reveal any differences in terms of learners' accurate use of the target structure: modal verbs and conditionals. Similarly, a targeted measure of accuracy (use of past forms for copula, regular and irregular verbs) was found in Kawauchi (2005) study and it is likely that this measure contributed to the finding that only low proficiency learners benefited in terms of accuracy with guided planning.

Lastly, although the relationships between CAF measures and task performance have been tested, less work has been done on analyzing L2 Chinese written texts using objective measures, such as T-unit based measures. In addition, the use of T-unit based measures in previous studies of L2 English might not guarantee their validity in measuring Chinese written production. As Robinson, Cadierno and Shirai (2009) noted "specific measures should be more sensitive to conception, task complexity, and its linguistic demands than general measures" (p. 550). The field of SLA is in need of valid and objective measure(s) or measurement units for L2 Chinese written texts.

CHAPTER 3 METHODOLOGY

3.1 Participants

A total number of 67 CFL learners from an intensive study-abroad program in Beijing, China initially signed up for the study. However, the data from 3 participants were discarded because they didn't complete the second writing session. This left 64 participants whose data were included in the analyses. The students' ages ranged from 21 to 37, with an average age of 24. There were 25 male and 39 female students respectively. On a background questionnaire, participants reported their native languages as English (n = 15), Japanese (n = 14), Korean (n = 12), Finnish (n = 6), Swedish (n = 5), Dutch (n = 4), Polish (n = 2), Thai (n = 2), Arabic (n = 2), Spanish (n = 1), and Portuguese (n = 1).

When they signed up to participate, students were enrolled in a Chinese course at the intermediate level (level 3 based on the university's summer program systems), which is equivalent to the ACTFL intermediate high or advanced low level. Before they came to the intensive program in China, all of the participants mentioned they had studied Chinese in high schools or at universities in their home countries. Regarding the amount of time they had spent in China, answers ranged from 7 months to 2 years, with the average being 14 months. About three quarters of the participants (77%) indicated that prior to the experiment, they had studied Chinese continuously for 1 to 2 years in the same intensive program in China. All participants indicated that they had had opportunities to be exposed to Chinese outside of class (including watching TV in

Chinese, listening to Chinese music, and reading in Chinese). The number of hours they reported spending on non-course-related Chinese learning each week ranged from 10 to 42 hours a week. On average, it was about 14 hours (*SD* = 4.26). All participants reported that they were engaged in conversations with Chinese native speakers between 30 minutes up to 4 hours, with an average of 1.5 hours on a daily basis (*SD* = 2.71).

The study-abroad program offers intensive Chinese classes covering all communicative language skills (reading, speaking, listening and writing), with a bi-weekly session of in-class writing on the topics of the reading texts that they have learned during the week. For each topic covered in the Chinese classes, learners participated in several contextualized vocabulary and grammar-focused activities. In general, the curriculum followed the input-practice-output progression. The instructor also confirmed that topic-comment constructions and topic chains had been covered in students' textbooks and were taught in the curriculum one semester prior to the study. Thus the participants didn't receive any special instruction on topic-comment constructions and topic chains for the study purposes. It was also confirmed by the instructor that the participants were able to use these constructions in their writings although the degree of accuracy varied among the students.

The participants came from four intact classes with the same instructor. Thus, the instruction was comparable for all participants. By the time the data were collected in the fall 2012 semester, the participants were enrolled in the Chinese courses at the advanced level, the highest level offered at the university. Their proficiency was further confirmed

by the instructor with the data of their untimed writings, which were used to "benchmark" the participants' writing proficiency in general. According to ACTFL Proficiency Guidelines for writing in Chinese (2012) published online, writers at the advanced level have the ability:

"to write routine informal and some formal correspondence, as well as narratives, descriptions, and summaries of a factual nature. They produce connected discourse of paragraph length and structure, and show good control of the most frequently used structures and generic vocabulary, allowing them to be understood by those unaccustomed to the writing of non-natives."

All participants were graded at the advanced or intermediate high levels based on their placement scores (M = 87/100, SD = 0.84). The decision to examine only advanced CFL learners' writing is based on the literature of formulaic language acquisition (Rott, 2004; Yorio, 1989). Processing and retrieving FSs in writing might be too demanding for lower proficiency learners. Furthermore, as the first study to report an interaction between planning and proficiency levels, Wigglesworth (1997) suggested that pre-task planning benefited only participants at a high proficiency level. She hypothesized that the low proficiency learners might use planning time to focus on content which resulted in a larger cognitive load. In contrast, advanced learners are capable of using the planning time to focus on either content or language forms at a manageable level.

The last rationale for including only advanced learners was drawn from studies on the acquisition of topic-comment structure and topic chains in L2 Chinese (Jin, 1994;

Yuan, 1995, Xiao, 2004). Although topic-comment constructions, base-generated topic and topic chains were introduced to learners early and frequently, learners didn't acquire and use these structures until they had reached a very advanced stage. Xiao (2004) summarized that zero pronouns and topic chains are the most difficult aspects of Chinese discourse for L2 learners to use. Since the current study focuses on comparing the length and ratio of topic-based syntactic features in L2 Chinese, it is reasonable to analyze only advanced CFL learners' writings because no or few instances of such syntactic features would likely be found in the written production of lower level learners.

One objective of this study was to use a repeated-measures approach to determine if L2 writers' use of FSs in writing would change according to planning condition. Participants were randomly assigned to one of four planning subgroups: LFP, CFP, LCFP and a minimal guidance group in two writing sessions (See Appendix II and III for essay topics). Four students were randomly selected for each planning condition from each of the four classes, constituting a subgroup of 16. Therefore, students from the same class wrote under different planning conditions. For instance, among the 16 students from class 1, four students were randomly assigned to the CFP group, four to the LFP group, four to the CLFP group, and another four to the minimal guidance group. Because of that, there were students from different classes under each planning condition.

Furthermore, to reduce the potential effect of topic influence, a second topic was selected that was considered comparable in nature and equally familiar to for the writers. Therefore, as an example, Student A from Class I wrote on the first topic in the LFP

condition and on the second topic, he/she would be instructed to write in either the CFP, the CLFP or the Minimal Guidance (MG) condition, and so on.

3.2 Experimental tasks

Participants were asked to write a reflective essay within a 45-minute limit based on the reading materials they had discussed during the week. In involvement load studies (e.g., Hulstijn & Laufer, 2001), vocabulary learning and retention are usually measured by means of immediate and delayed posttest, respectively. Kim (2011) used a timed writing task to measure new vocabulary learning and retention with regard to the "involvement load" it poses on learners.

The validity of using timed writing as a type of non- reciprocal task is supported by R. Ellis (2001, p. 49-50) who noted that "tasks can involve varying degrees of reciprocity" and "non- reciprocal tasks allow for the input to be scripted in such a way that it contains particular linguistic features" of which the learners' acquisition can be tested. "Reciprocity", in R. Ellis' terms, involved a two-way information exchange (usually between a speaker and a listener); while "non- reciprocal tasks" were that requiring only a one-way flow of information.

In fact, the writing task embodies every characteristic of a typical "task" within the task-based language teaching (TBLT) framework as defined by Skehan (1998, p. 95). A task is an activity in which "meaning is primary; there is a goal to be accomplished; the task is outcome evaluated; and there is a real-world relationship." In a similar vein,

Samuda and Bygate (2008) defined a task as "a holistic activity which engages language use in order to achieve some non-linguistic outcome while meeting a linguistic challenge, with the overall aim of promoting language learning, through process or product or both" (p. 69). The timed writing task also represents what Bygate, Skehan, and Swain (2001) defined as "a focused, well-defined activity, relatable to pedagogic decision making, which requires learners to use language, with emphasis on meaning, to attain an objective, and which elicits data which may be the basis for research" (p. 12). Similar timed writing tasks have been employed in previous studies by Kuiken and Vedder (2008) and Ong and Zhang (2010, 2013).

Students in the study were asked to read the instructions and prompts before they started to write. Students were reminded that the purpose of the reflective essay was to help them reflect critically on something they had read, learned, observed, felt, or experienced. The goal was to encourage them to express their feelings and to share their personal experiences, views, and thoughts on the topic.

3.3 Instruments

Background questionnaire. The background questionnaire, which was completed by the participants after the second writing session, was designed to elicit information regarding their previous study of Chinese (including how long they had studied Chinese and in what contexts), current uses of Chinese, length of stay in China as well as basic

bio data such as age, gender, and native language(s). A full version of the background questionnaire is provided in Appendix I.

The target FSs. A total of 30 target formulaic sequences of different categories were taken from two reading texts of the participants' textbook *Boya Chinese Intermediate* (Li, 2005), which was used by the intensive Chinese program at the university in China. All FSs were listed as useful phrases in the vocabulary and grammar sections of the textbook. The researcher and the instructor (with a background in linguistics) agreed upon the "formulaicity" of the target sequences and judged whether each sequence was a multi-word collocation (17), a phrasal frame (6) or a sentential connective (7). A complete list of the 30 target FSs is presented in Table 3.1. Native speakers' intuition has been argued as "an alternative technique of identifying lexicalized language in a given corpus," especially native speakers whose "intuition is shaped by professional experience" (Foster, 2001, p. 81-82).

Vocabulary Knowledge Scale (VKS). The thirty target FSs were presented with a Vocabulary Knowledge Scale (VKS) as an independent measure of the participants' explicit knowledge of the FSs. In the current study, Zimmerman's (1997) four point VKS was modified (See Appendix Table 1 for the original Chinese version and English translation). The original VKS was presented in Chinese to the participants without English translation. A mark in column A means "I don't know this phrase", column B "I know this phrase but I don't know how to use it", column C "I know this phrase and I can translate it" and column D "I can make a sentence with the phrase." Participants were

asked to indicate how well they knew each phrase on the VKS by placing a cross under the corresponding columns.

Table 3.1 *Categorization of formulaic sequences*

Multi-word collocation n = 17	Phrasal frames n = 6	*Categorization of Formulaic Sequences* Sentential connectives (n = 7)
1 家喻户晓 a household name	18 对……而言 means…to sb/sth	24 哪怕 …也 even if…still…
2 独具特色 unique and peculiar	19 有…….之称 has the title /name of	25 不仅...也 not only…, but…
3 世井文化 urban bourgeois culture	20 类似于 be similar to	26 无论…都 no matter…still…
4 侃侃而谈 to talk non-stop	21 迁怒于 vent one's anger on sb/sth	27…也好…也好 no matter (this) …and (that)…
5 温文尔雅 refine and cultivated	22…居多 the majority is…	28 是…还是… no matte…or…
6 精明能干 astute and shrewd	23 说不上… cannot tell whether…	29 若…就… if…then…
7 熙熙攘攘 hustle and bustle		30 一…就… once…then
8 鱼龙混杂 mixture of (good and bad)		
9 无伤大雅 without bothering sb/sth		
10 有点儿 a little bit		
11 满不在乎 do not care at all		
12 由它去 let it be		
13 闲得无聊 feel bored being idle		
14 有说有笑 talking and laughing		
15 凑热闹 come along for the fun		
16 自由自在 carefree and unstrained		
17 悠然遐想 to think leisurely and be lost in reverie		

Topics for writing tasks. Topic familiarity has been argued to affect attentional allocation in task performance (e.g., Nassaji, 2002). In order to minimize topic influence and examine whether the planning effects can be generalized to a similar topic, students writings were collected for two writing sessions with two similar topics that were comparable in nature and equally familiar to the writers (see Appendices III and IV for detailed writing instructions). Topic 1 asked students to write about a familiar city or place in the world, and topic 2 was about their preferred lifestyle or a habit. Even though participants had not read about the topics in their classes during the preceding week, both topics were believed to be general in nature, ensuring that all students had a similar level of familiarity and experience.

The participants also received the planning sheets (see Appendix IV for Topic 1, and Appendix V for Topic 2), designed specifically for each planning condition. The planning sheets guided participants in planning the content and/or the vocabulary based on the focus of planning. The decision to provide detailed planning sheets was based on Sangarun's (2001) pilot studies where she failed to observe participants' writing down sufficient words or shaping their discourse structure as expected. She concluded that the failure was due to inadequate guidance in the instructions.

3.4 Planning conditions

Regarding the length of pre-writing planning time, Mehnert (1998) suggested that only when at least a 10- minute planning was provided, were there measurable effects on

all three aspects of language use: fluency, complexity, and accuracy. Following previous studies (Ellis & Yuan, 2003, 2004; Foster & Skehan, 1996, 1999; Kawauchi, 2005; Ortega, 1999), the length of planning was set at 10 minutes.

As shown in Table 3.2, the language focused planning (LFP) group read a list of target FSs in Chinese, including word frame structures, phrasal frames, and sentence-level frame structures selected from the reading texts of the week. Participants were instructed to include as many target FSs as possible in their writing.

Table 3.2 *Treatment and task prompts of the different planning conditions*

Subgroups	LFP	CFP	LCFP	MG
Sample size	16	16	16	16
Treatment	Read a list of the target FSs selected from the reading texts of the previous week	Read elicitation questions with the target FSs embedded in the questions	Read both the elicitation questions (without the embedded FSs) and the list of the target FSs; FSs on the list were ordered in response to the elicitation questions	No specific guidance for planning; but students were allowed to plan for 10 minutes in the way they liked

Notes: CFP= content-focused planning
LFP= language-focused planning
CLFP=content- and language-focused planning
MG= minimal guidance
FS= formulaic sequence

For the content focused planning (CFP) group, participants read five elicitation questions which were based on the reading texts the participants had just learned and were crucial for planning the content. These questions also aimed to elicit ideas and examples that were likely to be discussed in the essay. Students in the CFP group were instructed to write their essays based on the elicitation questions with the embedded

target FSs, so they were provided with the same amount of exposure to the target FSs, although they might not have been aware of the usefulness of the FSs.

Participants in the content/language focused planning (CLFP) group were given both the elicitation questions (without the embedded FSs) as well as the list of the target FSs. It should be noted that FSs were ordered and presented corresponding to the five elicitation questions for the CLFP condition to facilitate answering the elicitation questions. The underlying assumption was that learners in the CLFP condition would be required to attend to both organization of ideas as well as linguistic forms, which in Robinson's words "involves greater mental and communicative effort, depth of processing, and so greater attentional and memory resource allocation to input" (2003, p. 54) compared to the other planning conditions. For all planning conditions, students were advised that they would have to write without their planning sheets.

3.5 Procedures for data collection

The researcher visited a total of four intact CSL classes and introduced the project to the students in the fall 2012 semester. Participants were asked to sign a consent form if they were willing to participate in the study. Before each treatment, the students also completed the background questionnaires. To ensure participants were at a similar level of proficiency, an untimed written task was completed a week before the timed writing task. The untimed writing samples were scored by the instructor holistically. The mean

score was 87 (of 100) and the data showed minimal variation among the students (*SD* = .17), indicating the writing proficiency was comparable between all groups of students.

Data from the timed writing assignments were collected from the same cohort of students during the in-class writing session. All participants participated in the two writing sessions with different topics in order to minimize topic influence, but as stated earlier, the type of planning condition was different for each participant's writing session topic. Participants in the planning conditions (including the minimal guidance group) planned for 10 minutes before they started to write. Students in different planning groups received specific guidance provided on the planning sheets; however, students in the minimal guidance (control) group were only given the topics. When the planning time was over, the planning sheets were collected and all students were asked to write a timed essay within 40 minutes. They were instructed that their writings would be judged based on content, length, language use, and grammar. There was a one week interval between the two writing sessions. The VKS was carried out a week before the first writing session and within a week after the second writing session.

The potential influence of the VKS on the students' writing was minimized in three ways: First, all the target FSs on the VKS were taken from the vocabulary list of the reading texts in the textbooks used in classroom instruction during the week of the reading sessions, so the participants had already been exposed to them. Second, the VKS was carried out a week before the first writing session. Third, there were both target and non-target FSs presented on the planning sheets during the planning stage. All words had

been introduced (with equal importance) during class sessions and the participants' attention was not directed only to the target ones.

Data collection was carried out during regular in-class writing sessions throughout the semester over a period of 16 weeks. Students from four intact classes participated in two writing sessions under different planning conditions. For example, Student A from Class I wrote on the first topic (A city or place) in the LFP condition and on the second topic (A lifestyle or habit) in the CFP condition, and so on. The design aims to test the effects of planning conditions by having the learners write on two different yet comparable topics. Within a week of the second writing session, participants were asked to complete the VKS as a post-test. Students were not informed about the objectives of the study until the end of the semester.

3.6 Data coding and scoring procedures

It should be noted what was measured in the current study was learners use and accurate use of already-learned vocabulary which were introduced to the learners during the week prior to their writing task. In other words, the target FSs were not new to the learners. For statistical analysis, a ratio was computed by dividing the number of correct FSs in each writing piece by the total number of FSs provided on the planning sheet in order to capture learners' uptake of the target FSs.

The accurate use of the target FSs was determined both in terms of the grammatical and pragmatic appropriateness of the FSs in the writings. To help examine

errors, an error classification scheme (Appendix VI) was developed based on Fan and Bai's (2007) work on identifying interlanguage errors for L2 Chinese. The researcher and the Chinese course instructor who had a Ph.D. in Chinese linguistics rated and coded all instances of FSs, Terminal Topic-Comment Unit (TTCU), error-free TTCU, topic chains (if there were any) and all the CAF measures (see Table 3.3 for a complete list of CAF measures). The interrater reliability ranged between .84 and .97 after independent coding. Any disagreements were resolved later by discussion.

Scoring on the VKS was relatively simple with 0 point assigned to column A (I don't know this phrase), 1 point to column B (I know this phrase but I don't know how to use it), 2 points to column C (I know this phrase and I can translate it) and 3 points to column D (I can make a sentence with the phrase). The highest possible individual score on the VKS was 90, based on 30 items. It should be noted that the two raters also checked the accuracy of the participants' translations (column C) and sentences (column D) when they marked in column C or D. Points were given only when the translations and FSs produced within sentences were correct. The two administrations of the VKS (pre-writing and post-writing) were scored separately.

For qualitative analysis, a sentence was deemed correct even if the formulaic sequence contained character errors. This decision was made because the Chinese language has few or no morphological changes that involve inflectional categories such as tense, aspect, mood, number, gender or case. As long as the sequence was used in a meaningful context with appropriate semantic and pragmatic functions, it was counted as

correct use. When students showed an intention to use the target FSs, but they were only partially correct (either syntactically correct but semantically inappropriate, or semantically appropriate but syntactically incorrect), the use was deemed inaccurate.

For instance, example (11), taken from the current study, represents an inappropriate use of the target FS.

(11) *生气也不能对别人迁怒于。

Shēngqì yě bù néng duì biéren qiānnù yú.

*Even if you're mad you can't to others be angry with.

In Example (11), the target FS was a verb + prepositional phrase 迁怒于*qiānnù yú* (be angry with…). Grammatically, the objects required by the structure could only be put in the fillable slots following the preposition in the FS, but not in front of it * 对别人迁怒于*duì biéren qiānnù yú* (to others be angry with). So in Example (11) even though the FS 迁怒于 *qiānnù yú* (be angry with…) was used correctly, the resulting syntactic structure was inappropriate.

3.7 Summary of the variables

To sum up, the independent variable investigated in the study was planning condition which was operationalized in four ways: content focused planning (CFP), language focused planning (LFP), content and language focused planning (CLFP) and minimal guidance planning (MG). Another independent variable was topic, the inclusion

of which allowed us to explore whether planning effects could be generalized to another topic albeit comparable under a similar task.

The fact that Chinese is typologically a topic-prominent (TP) language poses great difficulty in measuring the syntactic complexity of L2 Chinese written production, especially by advanced learners. Research shows that the language of advanced L2 Chinese learners exhibits a wide range of syntactic variation, including TP constructions and topic chains which are representative of the language of native Chinese speakers (Jin, 1994; Xiao, 2004; Yuan, 1995).

To explore valid measurement units of syntactic complexity, accuracy and fluency for L2 Chinese learners' written language, a pilot study was conducted based on Jin (1994, 2007) and Yuan (2009, 2010) to explore T-unit based and TTCU-based writing indices, including TTCU-based accuracy measures (total number of error free TTCUs, the ratio between the number of error free TTCUs and the total number of TTCUs), TTCU-based fluency measures (total number of words, total number of TTCUs) and TTCU-based complexity measures (mean length of TTCU, mean length of error free TTCU). Results from the pilot study indicated that TTCU-based units, but not T-unit based measures, were able to capture differences in terms of syntactic complexity for advanced learners from different L1 backgrounds. In other words, complexity measures were statistical significant only with TTCU-based units.

Table 3.3 *Summary of dependent and independent variables*

Independent variables	Meaning	Operationalization
Planning condition	LFP, CFP, CLFP, MG	based on teacher-driven focus on language or content
Topic		Session1: A city/place Session2: A lifestyle or habit
Dependent variables		
Complexity		
Mean length TTCU	average number of words per TTCU	total words in TTCU/total number of TTCUs
Clause per TTCU	average number of clauses per TTCU	total number of clauses/ total number of TTCUs
Accuracy		
# EFTTCU	total number of error-free TTCUs	
Mean length EFTTCU	average number of words per EFTTCU	total words in EFTTCU/total number of EFTTCUs
Fluency		
#words	total number of words	
FS use		
#FSs	number of occurrences of the target FSs	
FS uptake	accuracy ratio of target FSs	the number of correct FSs/the total number of FSs on the planning sheets
# Non-target FS	number of occurrences of the non-target FSs	
# Combined use		target FSs + non-target FSs occurrences

The battery of dependent variables is summarized in Table 3.3, including two TTCU-based complexity measures (mean length of TTCU, mean number of clauses per TTCU), two accuracy measures (total number of error free TTCUs, mean length of error free TTCUs), and one fluency measure (total number of words). The decision to include measures for both length and subordination to account for complexity is supported by

Bardovi-Harlig and Bofman (1989) and Ortega (2003) who suggest that a combination of syntactic complexity measures better captures the differences across proficiency levels. So the mean number of clauses per TTCU is actually the ratio between zero anaphora clauses or topic chains and TTCU in Chinese, the use of which exhibits features of native speakers' language. In addition, the number of target FS occurrences, FS uptake, non-target FS occurrences and the combined use of target and non-target FSs were also employed as dependent variables.

3.8 Data Analysis

The analyses were conducted using SPSS v 19.0. First of all, descriptive statistics (means, ranges, and standard deviations) were reported for each dependent and independent variable, including frequency counts for the number of FS occurrences, the number of TTCUs, and EFTTCUs for all participants under both writing sessions.

Relevant to the research questions 1, a MANOVA was chosen for comparing the planning effects on the eight dependent variables. Instead of multiple ANOVAs, the merit of a MANOVA analysis is to minimize the chance for Type I error. In addition, a MANOVA allows us to examine the interactions between several dependent variables. In other words, it has greater power than ANOVA to detect whether groups differ along a combination of dimensions/variables.

For MANOVA analyses, the assumptions are multivariate normality (the dependent variables are normally distributed collectively within groups) and

homogeneity of covariance matrices (equal homogeneity for each dependent variable and the correlations between any two dependent variables is the same in all groups) (see Field, 2005, p. 592 for details). The assumption of multivariate normality cannot be tested in SPSS, so alternatively, we need to check the normality for each dependent variable. For the assumption of homogeneity of covariate matrices, the first step is to check the univariate equality of variances using Levene's test, and then Box's test to compare the covariances between groups.

For the main MANOVA analysis, there are four commonly used methods: Pillai's trace, Wilks's lambda, Hotelling's trace and Roy's root (Field, 2005, p. 594). All four tests are robust to violations of multivariate normality. When sample sizes are equal, the Pillai's trace is the most robust to violations of assumptions. In the present study, both Pillai's trace and Wilks's lambda are reported for MANOVA analyses.

The alpha level for MANOVA was set at .05. When significant results were found, univariate ANOVAs were carried out as a follow-up analysis. Additionally, post hoc multiple comparison tests (Tukey HSD) were applied to locate significant differences between groups. Type I error resulting from performing multiple ANOVAs was controlled using Tukey HSD in SPSS.

For group differences in terms of target and non-target FS use, a series of ANOVA analyses were performed since ANOVA" is robust to violations of its assumptions" (Field, 2005, p. 542).

In addition, descriptive statistics and gains were reported on all target FSs. Paired sample t-Tests were conducted to measure the difference between scores on pre- and post-task VKS.

Finally, interrelationships among FS measures and all CAF measures were captured using Pearson correlation analysis, with the alpha set at .05. A correlation matrix was provided for each planning condition. For measures of FS use, non-target FS use and combined target and non-target FS use, Spearman Rank-order correlation coefficients were reported.

CHAPTER 4 RESULTS

4.1 Introduction

The results of the statistical analyses regarding each research question are reported in this chapter. A battery of dependent variables was employed, including two complexity measures (mean length of TTCU, clauses per TTCU), two accuracy measures (total number of error free TTCUs, mean length of error free TTCU), one fluency measure (total number of words), and four FS use measures (total number of target FSs per essay, FS uptake, non-target FS use as well as combined use of target and non-target FSs). Planning effects were examined using a MANOVA and the interrelationships among all measures were reported in a correlation analysis.

The means and standard deviations for the CAF measures (average of both writing sessions) are displayed in Table 4.1. Both the assumption of equality of covariance matrices and assumption of homogeneity were met as indicated by Box's test ($p > .05$) and Levene's tests ($p > .05$). The results of the MANOVA and the subsequent ANOVAs are shown in Table 4.2. Results from the Tukey HSD post hoc multiple comparison tests are presented in Table 4.3. Table 4.4 summarizes the monotonic relationship across planning conditions. Concerning RQ 2, the results from a series of ANOVA analyses regarding the effects of planning condition on the use and uptake of target FSs are displayed in Table 4.5 – 4.7. Furthermore, correlation analyses which investigated the interrelation of the accuracy, fluency, complexity measures and FS use are displayed in Table 4.8 (with both Pearson's and Spearman's correlation coefficients).

Finally, descriptive statistics and gains on all target FSs were reported in Table 4.9. Finally, a list of non-target FS use was presented in Table 4.10.

4.2 Results concerning Research Question 1

Research Question 1: What are the effects of manipulating planning conditions, namely, language focused planning (LFP), content focused planning (CFP) and content and language focused planning (CLFP), on L2 Chinese learners' written products in terms of the complexity, accuracy and fluency (CAF) measures?

CAF measures were averaged with scores from both writing sessions. In Table 4.1, the means and standard deviations of CAF measures are presented. Regarding the complexity measure of clauses per TTCU, scores were the highest under the CLFP condition, followed, respectively, by the CFP condition, LFP, and the MG condition. For mean length TTCU, the CLFP outperformed the LFP group, which in turn, scored higher than the MG group, followed by the CFP group.

Accuracy was measured in two ways: total number of error-free TTCUs and mean length of error free TTCU. For the measures of number of EFTTCU, the highest scores were found in the LFP condition, followed by the CLFP group, and similar scores were found between the CFP and the MG condition. For mean length EFTTCU, the LFP group again outperformed the other groups. Mean scores under the LFP condition were the highest, followed, respectively, by the CLFP condition, the CFP condition, and the MG condition.

The descriptive statistics of the fluency measures suggested little variation among the planning conditions in total number of words. Learners in all planning groups produced a comparable number of Chinese characters (text length), ranging from the shortest essay with a total of 186 characters to the longest one with 392 characters. Nevertheless, the CFP and the MG group actually outscored the LFP and the CLFP groups in the number of words produced.

Table 4.1 *Descriptive statistics for CAF measures*

Measures		CFP	LFP	CLFP	MG
Complexity					
clauses per TTCU	***M***	1.6	1.5	2.0	1.2
	SD	1.4	2.0	2.2	2.1
mean length TTCU	***M***	14.2	15.9	18.9	14.7
	SD	1.8	2.3	2.7	1.3
Accuracy					
# EFTTCU	***M***	9.5	14	11	9
	SD	3.1	4.1	4.8	4.2
mean length EFTTCU	***M***	14.2	16.6	15	13
	SD	4.1	3.2	3.5	3.4
Fluency					
# words	***M***	234	231	224	234
	SD	3.0	3.3	2.0	4.3

Notes: CFP = content focused planning; LFP = language focused planning; CLFP = content and language focused planning; #words = number of words; EFTTCU = error-free TTCU; mean length EFTTCU = average number of words per EFTTCU; clause per TTCU = average number of clauses per TTCU; mean length TTCU = average number of words per TTCU

Table 4.2 *Summary of one-way ANOVAs on CAF measures*

Measures		*df*	*F*	*p*
Complexity	clauses per TTCU	3	7.42	.026
	mean length TTCU	3	17.50	.000
Accuracy	# EFTTCU	3	11.42	.014
	mean length EFTTCU	3	9.59	.020
Fluency	#words	3	2.89	.340

MANOVA analysis indicated there was a statistical significant difference among the planning conditions with combined effects on complexity and accuracy. Follow-up one-way ANOVAs (Table 4.2) further identified the differences.

Table 4.3 *Results of post hoc comparison tests*

Measures		Group comparisons	*t*	Sig*
Complexity	clauses per TTCU	CLFP > LFP	3.71	.030
		CLFP > MG	5.25	.018
		CFP> MG	8.25	.014
	mean length TTCU	CLFP > CFP	13.00	.000
		CLFP > MG	11.40	.004
Accuracy	# EFTTCU	LFP> CFP	8.44	.016
		LFP > MG	7.05	.017
	mean length EFTTCU	LFP> MG	9.11	.022

**Note:* Only significant results are reported.

Results from the post hoc comparisons (Table 4.3) indicated that for the complexity measure of clauses per TTCU, there were statistical differences between the CLFP and LFP groups. In addition, both CLFP and CFP groups outscored the MG group on clauses per TTCU significantly. For the measure of mean length TTCU, the CLFP group also performed better than the CFP group and the MG group.

For the two accuracy measures, post-hoc tests indicated that the LFP group outscored the CFP and MG groups on total number of EFTTCU. In addition, the LFP group also performed significantly better than the MG group on mean length EFTTCU.

Finally, results from the post hoc comparisons indicated that the fluency measure didn't yield significant results for any pairwise comparison. In other words, learners in

the LFP, CFP and CLFP groups did not necessarily write longer essays than those in the MG group. Table 4.4 summarizes the monotonic relationship across planning conditions.

Table 4.4 *Summary of the monotonic relationship across planning conditions*

Measures	Monotonic relationship	Significant differences
Complexity		
Clauses per TTCU	CLFP > CFP > LFP > MG	CLFP > LFP CLFP > MG CFP > MG
mean length TTCU	CLFP > LFP > CFP > MG	CLFP > CFP CLFP > MG
Accuracy		
# EFTTCU	LFP > CLFP > CFP > MG	LFP > CFP LFP > MG
mean length EFTTCU	LFP > CLFP > CFP > MG	LFP > MG
Fluency		
words	MG = CFP > LFP > CLFP	n.s.

4.3 Results concerning Research Question 2

Research Question 2: What are the effects of manipulating planning conditions on L2 Chinese learners' correct use of the target FSs in writing? Will the participants in the LFP group or the CLFP group (with a focus on both content and language) be prompted to use more target FSs than the participants in the other two groups? For non-target FS use, will there be significant difference across planning conditions?

The research questions prompted an explanation of the planning effects on target FS use and uptake (indicated by the ratio of correct FS use by the total number of target FSs). The primary assumption is that participants in the CLFP will perform as well as

those in the LFP group if they are capable of attending to both form and content during planning.

Means and standard deviations (in parentheses) for the number of target and non-target FS occurrences and the target FS uptake scores are reported in Table 4.5. In general, the performance scores were rather low, indicating learners' difficulty in retrieving and reproducing the target FSs in writing. Note that the data of target FS uptake by the minimal guidance group were excluded from the analysis since participants in the MG group were not provided with planning sheets (with target FSs).

Table 4.5 *Target and non-target FS use and FS uptake across planning conditions*

	LFP	CLFP	CFP	MG
# target FS	4.4	3.7	2.0	1.1
Session 1	(2.6)	(1.5)	(2.2)	(3.3)
# target FS	3.6	3.9	2.8	1.7
Session 2	(1.1)	(0.8)	(2.0)	(2.5)
# non-target FS	1.3	1.2	1.6	1.6
Session 1	(2.8)	(4.0)	(3.3)	(4.2)
# non-target FS	1.2	1.5	1.8	1.9
Session 2	(1.7)	(1.9)	(3.0)	(3.3)
FS uptake	0.24	0.23	0.20	n/a
Session 1	(1.3)	(0.7)	(1.2)	
FS uptake	0.22	0.19	0.14	n/a
Session 1	(0.8)	(1.4)	(1.7)	

Notes: #FSs: number of occurrences of the target FSs

FS uptake: the number of correct FSs/the total number of FSs per essay

Results from three ANOVA analyses (Table 4.6) indicated that there were significant differences between the target FS use and uptake across planning conditions. Furthermore, results from post hoc comparisons (Table 4.7) comparisons indicated that both LFP and CLFP groups performed significantly better than MG groups on target FS

use. The LFP and CLFP groups also outscored the CFP group in terms of FS uptake. No significant differences were found in terms of non-target FS use across planning conditions.

Table 4.6 *Summary of one-way ANOVAs on FS use*

Measures	*df*	*F*	*p*
# target FS	3	8.22	.013
# non-target FS	3	3.50	.052
FS uptake	2	6.74	.034

Table 4.7 *Results of post hoc comparison tests*

Measures	Group comparisons	*t*	Sig*
# target FS	LFP > CFP	5.25	.030
	LFP > MG	11.44	.006
	CLFP > MG	9.89	.011
FS uptake	LFP> CFP	3.72	.042
	CLFP> CFP	3.71	.042

**Note:* Only significant results are reported.

4.4 Results concerning Research Question 3

Research Question 3: Are there trade-offs between complexity and accuracy measures? What are the relationships between the CAF measures and learners' usage of the target FS under each of the planning conditions?

The last research question aims to explore the interrelations among the CAF measures, especially, whether there are trade-offs between complexity and accuracy measures. Tables 4.8 report correlations (Pearson's *r*) among CAF measures. For

measures of target and non-target FS occurrence and combined FS use, Spearman's rank-order correlation coefficients ρ were calculated.

It was found out that five pairs of correlations reached the .05 level of significance. Significant correlations were found between the two accuracy measures ($r = .32$, $p < .05$), and between the two complexity measures ($r = .46$, $p < .01$).

Positive significant relationship was also found between total number of EFTTCU (accuracy) and mean length TTCU (complexity) ($r = .43$, $p < .01$). By examining the correlation coefficients between the complexity and accuracy measures, it seemed that they didn't fluctuate in opposite directions, indicating no observed trade-offs. Furthermore, no significant relationships were found between complexity and fluency measures, and between accuracy and fluency measures.

Table 4.8 *Correlation matrix for the dependent variables*

	F1	A1	A2	C1	C2	FS1	FS2
#words (F1)							
# EFTTCU (A1)	–.19						
mean length EFTTCU (A2)	–.02	.32*					
clauses per TTCU (C1)	.20	.25	–.15				
mean length TTCU (C2)	.06	.43**	.14	.46**			
#FS (FS1)	.17	.09	.07	.13	.22		
FS uptake (FS2)	.13	.18	.24	.31*	.25	.21	
Non-target FS	.10	.14	.18	.22	–.04	.11	–.12
FS combined	-.14	.11	.13	.27*	.10	.16	.08

Note: * = p <.05; ** = p <.01

FS combined = target FSs + non-target FS occurrences

Finally, significant positive correlations were found between target FS uptake and clauses per TTCU (complexity) ($\rho = .31$; $p < .05$), and between target and non-target FS combined scores and clauses per TTCU (complexity) ($\rho = .27$; $p < .05$).

4.5 Data from the planning sheets

Planning sheets were collected at the end of the planning session. It turned out that most participants left their planning sheets blank, with only an occasional underlining or circling of the original texts. There were a few instances when participants wrote down words or phrases in their native languages, indicating they were planning or searching for words or phrases in particular, their meaning. The other type of potential planning activity was to write down words in Chinese. Usually in this case, participants seemed less sure about the spelling or the order of the Chinese characters, as these words in Chinese were often accompanied by cross-outs or apparent attempts to get them right, indicating monitoring behaviors. In the meantime, no clear traces were found for planning on content in the collected planning sheets. It is possible that participants were planning mostly in their heads.

4.6 FS scores on the VKS

Table 4.9 reports the differences between participants' pre-writing VKS and those on the post-writing VKS. On two instances, a negative sign in front of the number indicates a lower score on the post-test than that on the pre-test. Based on the differences between the post-writing and pre-writing VKS scores, it could be concluded that gains were mostly shown on the use of multi-word collocations, such as 鱼龙混杂 *yúlónghùnzá* (mixture of good and bad) (2.6) and 独具特色 *dújùtèsè* (unique and peculiar) (2.4), despite the fact that the gain on the multi-word collocation 熙熙攘攘 *xīxīrăngrăng*

(hustle and bustle) was negative (-0.2). At the same time, none of the gains on sentential connectives sentential connectives such as 一…就… *yī…jiù…* (once…then…) exceeded 1.6 on the VKS. For the sentential connective 也好…也好… *yě hǎo … yě hǎo* (no matter (this) …and (that)…), the gain was a negative 0.2.

Table 4.9 *Differences between participants' pre- and post-writing VKS scores*

Target FSs (n = 30)	Average scores on pre-writing VKS	Average scores on post-writing VKS	Gains
鱼龙混杂 mixture of (good and bad)	0.2	2.8	2.6
独具特色 unique and peculiar	0.6	3.0	2.4
无伤大雅 without bothering sb/sth	0.2	2.4	2.2
类似于 be similar to	0.6	2.8	2.2
凑热闹 come along for the fun	0.6	2.8	2.2
世井文化 urban bourgeois culture	0.6	2.6	2.0
迁怒于 vent one's anger on sb/sth	0.2	2.2	2.0
有…….之称 has the title /name of	1.0	3.0	2.0
温文尔雅 refine and cultivated	1.2	3.0	1.8
对……而言 means…to sb/sth	1.2	3.0	1.8
闲得无聊 feel bored being idle	0.6	2.2	1.6
说不上… cannot tell whether…	0.4	2.0	1.6
侃侃而谈 to talk non-stop	0.4	2.0	1.6
一…就… once…then…	0.6	2.2	1.6
有点儿 a little bit	1.2	2.8	1.6

Table 4.9 (cont'd)

…居多 the majority is…	1.4	2.8	1.4
悠然遐想 to think leisurely	0.2	1.6	1.4
哪怕 …也 even if…	0.4	1.8	1.4
无论…都 no matter…still…	0.6	2.0	1.4
若…就… if…then…	1.0	2.2	1.2
是…还是… no matte…or…	0.8	2.0	1.2
有说有笑 talking and laughing	1.0	2.4	1.2
由它去 let it be	1.0	2.2	1.2
不仅...也 not only…, but…	1.4	2.4	1.0
自由自在 carefree and unstrained	1.2	2.0	1.0
满不在乎 do not care at all	1.4	2.4	1.0
家喻户晓 a household name	1.0	1.6	0.6
熙熙攘攘 hustle and bustle	1.0	0.8	-0.2
…也好…也好 no matter (this) …and(that)…	0.6	0.4	-0.2

Results from paired sample *t*-Tests (between pre- and post-test VKS) indicated that gains on multi-word collocations were statistically significant ($t = 14.96$, $p < .05$). No significant differences were found in terms of gains on phrasal frames ($t = 6.15$, $p > .05$) and sentential connectives ($t = 5.38$, $p > .05$).

4.7 Non-target FS use

As discussed in 4.3, no significant differences were found in terms of non-target FS use across planning conditions. In other words, learners performed rather similarly using non-target FSs in both writing sessions regardless of the planning condition. A close examination on the non-target FS use revealed some interesting results. First, learners in the MG group seemed to resort to non-target FSs that were semantically comparable to the target ones, given the fact that they were not provided with the planning sheets during planning.

(14) 在大学路, 不仅可以喝各种各样的酒, 而且还可以享受多样化的文化演出.

Zài dàxué lù , bújǐn kěyǐ hē gèzhǒnggèyàng de jiǔ , érqiě hái kěyǐ xiǎngshòu duō yàng huà de wénhuà yǎnchū.

(You will not only be able to drink all kinds of wines at College Street, but also to enjoy a multitude of artistic performance.)

(15) 这里既有三越百货等历史悠久的日本百年老店, 又有许多世界名牌商品的专卖店.

Zhèlǐ jì yǒu sānyuè bǎihuò děng lìshǐ yōujiǔ de rìběn bǎinián lǎodiàn , yòu yǒu xǔduō shìjiè míngpái shāngpǐn de zhuānmàidiàn .

(There are not only renowned department stores such as Mitsukoshi with over 100 years of history, but also many boutiques for international brands.)

Examples (14) and (15) illustrated the use of replacement phrases 不仅…而

且…*bújǐn… érqiě…* (not only… but also…) and 即…又… *jì …yòu…* (not only… but also…) to the target one 不仅…也…*bújǐn…yě…*(not only… but also…). All three FSs were semantically and syntactically comparable and interchangeable without interfering the meaning of the original sentence.

Secondly, the range and types of FSs varied especially for the CFP and MG groups while participants were not given the planning sheets or no direct instructions about focus on the target FSs. Table 10 summarized all the non-target FSs use in the CFP and MG groups. A total of 22 multi-word collocations, 10 phrasal frames and 5 sentential connectives were reported. After checking with the class instructors, it was found out that the all of the non-target FSs produced by the learners from the CFP and MG groups could be located in the textbook *Boya Chinese* (Li, 2005).

Table 4.10 *List of non-target FS use (in the CFP and MG groups)*

Multi-word collocation n = 22	Phrasal frames n = 10	Sentential connectives n = 5
卖关子	从…到…	尽管…也 …
Keep others guessing	From…to…	Even though…also…
随处可见	依我看	即…又...
(can be) seen everywhere	In my view,	Not only…but also…
首屈一指	跟…似的	除了…也…
The greatest/leading	Similar to …	Except…also…
转祸为福	对…有兴趣	当…就…
Turn a curse for the blessing	Be interested in…	When…just…
秩序井然	出于…的考虑	有的人…有的人…
Orderly and organized	Out of the consideration for…	Some people…other people…
不声不响	非…不可	
Quietly (without fuss)	Won’t (work) without…	
鱼贯而入	相对地说	
Come in succession	Relatively speaking,	
说三道四	换句话说	
Gossip		

Table 4.10 (cont'd)

酒足饭饱 Filled up with wine and meal	In other words, 一般来说
大手大脚 Wasteful and liberal	In general, 总的来说
五花八门 A wide variety of	In conclusion,
老少皆宜 All age appropriate	
十里洋场 Metropolis infested with foreign adventurers (referring to the city Shanghai before 1949	
年富力强 Young and energetic	
乐此不疲 Without feeling tired	
最富盛名 With the best reputation	
与众不同 Unique and outstanding	
至今为止 So far	
和谐发展 Harmonious development	
香气四溢 Overflow aroma	
历史久远 With a long history	
万变不离其宗 Remaining the same in essence	

4.8 Error analysis of FSs

Furthermore, analysis on the use of the FSs in learners' written production revealed that certain errors were consistent. For instance, the most commonly observed error types related to target FSs are analyzed and are illustrated below in examples (16) – (19).

One type of error multi-word collocation was related to verb transitivity.

(16) *北京的普通市民喜欢侃侃而谈政治。

Běijīng de pǔtōng shìmín xǐhuan kǎnkǎnértán zhèngzhì.

**Ordinary Peking locals like talking politics non-stop.*

(17) *他满不在乎周围人吸烟。

Tā mǎnbùzàihu zhōuwéi rén xīyān.

**He doesn't care at all people around him smoke.*

In Example (16), 侃侃而谈 *kǎnkǎnértán* is an intransitive verb phrase. As such, the sentence became syntactically ungrammatical with the direct object 政治 *zhèngzhì* attached to the verb phrase. Likewise in Example (17), attaching an objective clause 周围人吸烟 *zhōuwéi rén xīyān* to the target FS 满不在乎 *mǎnbùzàihu* (which is an intransitive verb phrase) made the sentence grammatically incorrect.

(18) *生气也不能对别人迁怒于。

Shēngqì yě bù néng duì biéren qiānnù yú.

**Even if you're mad, you can't to others be angry with.*

In Example (18), the target FS 迁怒于*qiānnù yú* (be angry with) is a verb + prep phrase. Grammatically, the objects required by the structure can only be put in the fillable slots following the preposition 迁怒于别人*qiānnù yú biéren* (to others be angry with), but not in front of it. So in Example (18) even though the target FS was reproduced intact, the resulting syntactic structure was inaccurate.

(19) *无论是中午还是半夜, 你一直会觉得非常安全。

Wúlùn shì zhōngwǔ hái shì bànyè , nǐ yīzhí huì juéde fēicháng ān'quán.

* *No matter it is noon or midnight, you will always feel very safe.* (*still* is needed to complete the sentential connective in Chinese).

(20) 哪怕我在别的城市, 我想着在纽约的生活.

Nǎpà wǒ zài bié de chéngshì , wǒ xiǎngzhe zài niǔyuē de shēnghuó.

**Even if I'm in another city, I miss life in New York.* (*also* is needed to complete the sentential connective in Chinese).

In Example (19), the second part 都 *dōu* (still) in the target sentential connective 无论…都… *wúlùn … dōu.*… (no matter… still…) was missing. In Chinese, sentential connectives often appear in pairs, so missing 都 *dōu* (still) resulted in an incorrect syntactic structure. Likewise in Example (20), 也 *yě* (also) is obligatory in the sentential connective 哪怕…也… *nǎpà… yě* (even if…also…), so the missing component 也 *yě* (also) made the sentence grammatically incorrect.

To summarize, qualitative analyses of the FS related errors revealed that errors are specific to each type of FSs, which include verb transitivity errors, FS induced incorrect word order, and missing part(s) in sentential connectives. It was postulated that error types were influenced by the structural features of FSs. The error types were mostly tied to syntactic inaccuracy (such as verb transitivity errors, FS induced incorrect word order and missing part(s) in sentential connectives), which support Zhan's (2012) and Wang's (2013) findings of errors concerning the immediate constituents preceding or following the FSs.

4.9 Syntactic complexity features in L2 Chinese written output

L2 Chinese learners' written data in the study also exhibited certain syntactic features that are comparable to those of the native speakers' language, including the use of zero pronouns/anaphors, coreferential relationship and topic chains (Jin, 1994; Yuan, 1995, Xiao, 2004). Below are three examples from the written texts of participnats in the CLFP condition.

(21) 我是在赫尔辛基出生的，在赫尔辛基长大的。/我也在赫尔辛基上大学。

Wǒ shì zài hè ěr xīn jī chūshēng de, zài hè ěr xīn jī zhǎngdà de. Wǒ yě zài hè ěr xīn jī shàng dàxué.

I was in Helsinki born de, Ø in Helsinki grow up de. I also in Helsinki attend college.

(I was born and grew up in Helsinki. I also attended university in Helsinki.)

There are 28 characters in example (21), with 3 T-units and 2 TTCU (segmented with "/" in the example). The first 2 T-units form a topic chain in Chinese with a zero anaphor. The mean length of TTCU is 14.

(22) 神奈川曾经是有将军镇守的地方，在 1900 年代曾是国际贸易的源头，当时的"小田原城"是今天家喻户晓的名胜古迹之一，总之是独具特色的地方。

Shén nài chuān céng jīng shì yǒu jiāngjun zhèn shǒu de dìfāng, zài 1900 niándài céng shì guójì màoyì de yuántóu, dāngshí de "xiǎo tián yuán chéng "shì jīntiān jiā yù hù xiǎo de míng shèng gǔ jì zhī yī, zǒngzhī shì dú jù tè sè de dìfāng.

Kanagawa once was have general guard de place, Ø at 1900 era is international

business –de origin, Ø that time de “small rural village” is today known-to-all families famous old ritual sites of one, Ø sum up is unique characteristics -de place.

(Kanagawa was once guarded by generals. In 1900 it is the place of origin for international business. A “small rural village” at that time is nowadays one of the famous places of interests. To sum up, it is a place of unique characteristics.)

Example (22) has a total of 54 Chinese characters, with 4 T-units. But this long sentence contains only one TTCU with 3 occurrences of zero anaphora. The mean length of TTCU is 54.

Another interesting observation of the data is that there were a number of instances of double-link topic chains, such as example (13) by a Korean student in the CLFP group from writing session 1.

(23) 现在去汝矣岛公园玩的时候，可以看到各式各样的餐厅，咖啡厅和酒摊儿，在夜晚时特别热闹。

Xiànzài qù rǔyǐdǎo gōngyuán wán de shíhòu, kěyǐ kàn dào gè shì gè yàng de canting, kāfēi tīng hé jiǔ tānr, zài yèwǎn shí tèbié rènào.

Now Ø go Ruyi island park play -de time, Ø may see -Comp various -de restaurants, cafés and pubs, Ø at night time especially lively.

(If you go and play in Ruri island park now, you may see various restaurants, cafés and pubs, which are especially lively at night.)

Although there is no overt subject (null subject) in (23), it's not hard for native speakers of Chinese to decide on the subject "you." Thus the first clause is a clause with zero pronoun. Also, there are two instances of zero anaphora and two coreferential relationships, making this sentence a double-link topic chain. The controlling topic for the second clause is "you", and "restaurants, cafés and pubs" have control over the third clause. All in all, evidence of longer TTCUs (measured as mean length TTCU), to some extent, was observed in the L2 Chinese written data in the study, representing syntactic features of native speakers of Chinese' writing.

CHAPTER 5 DISCUSSION

5.1 Introduction

Results of the current study, to a large extent, supported the hypotheses and were consistent with the findings from Kuiken and Vedder's (2007) and Ong and Zhang's (2010) studies on planning influence on the complexity, accuracy, and fluency measures of L2 written output. The discussion centers on the rationale of each CAF measure. By examining closely the nature of CAF measures, it seems that even Skehan's Limited Attentional Capacity model argues there are trade-offs between complexity and accuracy, results in the current study indicate task condition/instruction could affect complexity and accuracy simultaneously, that is, at the surface level, learners seemed to be able to attend to both linguistic accuracy and syntactical complexity during pre-writing planning.

5.2 Effects of the planning conditions on CAF measures

5.2.1 Complexity

Foster and Skehan (1999) argued that content focused planning would lead to greater complexity while language focused planning would lead to higher accuracy. So there would be trade-offs between accuracy and complexity. According to Foster and Skehan (1996, 1999), both accuracy and complexity are primarily related to L2 knowledge representation.

In this study, participants in the CLFP group were directed to focus on both the target FSs and the questions highlighting the content of the article they read. It turned out

that participants from both the CLFP and LFP groups outperformed the CFP group in accuracy and complexity; nevertheless, at the same time, there were no statistically significant differences in complexity measures between the LFP and CLFP groups. Furthermore, participants from the LFP and CLFP groups performed similarly on accuracy measures as well. In other words, when participants were prompted to focus either on language or both language and content, they prioritized the accuracy of their productions (as indicated by the differences in terms of accuracy between LFP and CLFP), and then the linguistic complexity of the text.

It's not hard for us to take Foster and Skehan's claim that unlike fluency, both complexity and accuracy are more relevant to L2 knowledge representation, but the problem is whether there will be trade-offs. Results of the current study indicated that even though scores on the complexity measures did not always mirror those on the accuracy measures, it was clear that participants did not sacrifice accuracy for better complexity when prompted to plan content, nor did they sacrifice complexity for accuracy when prompted to plan language form. The obvious evidence is that participants from both the LFP and CLFP groups did not differ in complexity, but they both outperformed the CFP group (and the minimal guidance group) in terms of complexity measures.

Findings should be interpreted with some caution since four types of planning conditions (LFP, CFP, CLFP and minimal guidance) were addressed here; a conclusion about general complexity is challenging. The nonlinearity between complexity measures

in L2 acquisition was also discussed in Norris and Ortega's (2009) study. They noted that coordination might be most helpful for distinguishing lower-level learners, subordination for intermediate level, and phrasal complexity for advanced learners. Similarly, Pallotti (2009) suggested that depending on the measurement unit, syntactic complexity (e.g., coordination and subordination) would differ drastically from phrasal complexity (average clause length) for measuring L2.

According to Kuiken et al., complexity is the "most tricky, elusive and hardly captured" phenomenon (2008, p. 91). Given the fact that complexity is multi-faceted, the main conclusion seems to be that syntactic complexity did not suffer when learners' attention was directed to both language and content. At the same time, when their attention was directed to plan only content, participnats did not necessarily seize the opportunity to prioritize syntactic complexity. In the following section 5.4, the interrelationship among all CAF measures and between the CAF measures and FS use will be discussed in detail.

5.2.2 Accuracy

As discussed earlier, it seems that no matter what the planning conditions were, students always focused on linguistic accuracy in the first place. For advanced L2 learners in the current study, the influence of pre-writing planning on language held during L2 written production, resulting in the greatest accuracy observed in the LFP group, following by the CLFP, CFP and MG groups. The results were partly in line with

Kuiken and Vedder's (2007, 2008), Michel et al.'s (2007) and Mochizuki and Ortega's (2008) studies. It seems that when given the opportunity to focus on the target FSs (as in both LFP and CLFP groups), the participants were more likely to write more accurately as measured by two general accuracy measures. In contrast, participants in the CFP and MG conditions did not seem to have attended to the accuracy of the linguistic forms they produced as much as when they performed in the LFP and CLFP conditions.

The role of language focused planning in promoting accuracy was confirmed by Wendel (1997) who suggested providing task instructions to attend to form immediately prior to performance may be the only means to increase accuracy. The reason is that accuracy of L2 performance results largely from on-line processes (i.e., monitoring) while pre-task planning is a strategic type of planning that engages off-line processes. If there is no directed attention to form as required by the task, off-line planning effects do not transfer automatically to benefit on-line processes.

With respect to Skehan's (2009) model, accuracy is the "consequence of attention being available for Formulator operations" (p. 518) which is quite similar to Wendel's argument of the on-line nature of accurate performance. Tavakoli and Foster (2008) and Foster and Tavakoli (2009) found that for a complex task (such as one which requires both foreground and background information), the increase in accuracy was the result of the task design feature, with the structure of the task affecting performance accuracy negatively. In the meantime, syntactic complexity in L2 production was influenced by learners' manipulation of information. What's more, the baseline performance from both

studies suggested that the participants prioritized attention either to accuracy or complexity, which is not a natural thing to do unless the task condition/instruction directed them to do so.

The possible simultaneous task effect on accuracy and complexity is compatible with the predictions of Robinson's Cognition Hypothesis that both complexity and accuracy will increase with more complex tasks, and task performance is not constrained by attentional limitations. It is very likely that the effects of planning on accuracy in the study resulted from the fact that both CFP and CLFP planning conditions were effective in directing the learners to focus on accuracy, and thus reducing the processing loads in written production so that the learners were able to monitor the written production involving syntactic and clausal complexity to some extent.

Tavakoli and Foster (2008), Foster and Tavakoli (2009) and Skehan (2009) attributed trade-off effects to task manipulation: How and how detailed are the task instructions given to the learners? In other words, trade-offs could be attributed to task instruction rather than "task complexity" per se. Regarding limited attentional capacity, Foster and Skehan (1999) argued that the competition between complexity and accuracy could be alleviated with input from teacher's instruction and planning manipulation (e.g., the availability of target FSs and facilitative questions for essay content development). Thus, when complexity and accuracy may look like they perform in parallel and simultaneously, the underlying mechanism for each may work separately but additively.

In the current study, the differences in terms of accuracy and complexity observed between the LFP/CLFP and CFP/MG groups suggested a task manipulation effect. Further evidence about the correlations between accuracy and complexity as well as those among the six CAF measures will be discussed in the following section.

5.2.3 Fluency

There is no doubt in either the Cognition Hypothesis or the Limited Attentional resources model that fluency, unlike complexity and accuracy, is more of a performance measure suffering from high processing effort (Levelt, 1989). Skehan and Foster (1999) maintained that fluency is a different construct than accuracy and complexity. In a similar vein, Towell (2012) asserted that unlike complexity and accuracy, fluency concerns with proceduralization of linguistic processes in the model of CAF measure constructs he proposed.

In the current study, the fluency measure did not fluctuate significantly across planning condition. One possible explanation could be that the fluency measure has been shown in other studies to be effective at distinguishing between each of the Novice-Mid through Intermediate-High levels, while fluency was not the most effective at distinguishing among higher proficiency samples (Norris & Ortega, 2009) as is the case for the current study.

All participants were enrolled in the Chinese course at the advanced level, the highest level offered at the university by the time the data were collected. They had

already spent 1 to 2 years learning Chinese continuously in the same program. After checking with the course instructor, it was confirmed that these students made only a few morphological and syntactic errors in their writings, and they were quite comfortable in understanding and producing native-like sentences with topic-comment constructions and topic chains. With this understanding, it is not hard to expect that merely counting the number of words was not enough to capture the differences in fluency. For instance, Wolfe-Quintero, Inagaki and Kim (1998) suggested that number of words per T-unit and number of words per clause are the best measures of development in fluency. Actually, mean length TTCUs (number of words per TTCU) was used as a measure for complexity in the study. As Wolfe-Quintero, Inagaki and Kim also pointed out that mean length measures are best for capturing development differences in fluency, they are not necessarily robust to distinguish written fluency of learners at comparable proficiency levels.

5.3 Effects of planning on the use of target FSs

Participants from both the LFP and CLFP groups exceeded the CFP group and the MG group in terms of the number and accurate use of the target FSs. The results supported the role of pre-task planning in facilitating formulaic language acquisition by L2 learners (see the use of other pre-task activities such as vocabulary and grammar list in Foster, 2001 and brainstorming in Rott, 2009). In Robinson's words, learning from the

input is possible when “forms were made salient in the input” and there was “heightened attention to and memory for input” (2011, p. 48).

The central issue here is whether and how learners’ attention can be directed to the target linguistic forms during the pre-task planning stage. Unfortunately, the current research design does not permit a response to the question without other supportive data from a think-aloud protocol or stimulated recall. However, the rigorous comparisons among the four planning conditions in the study provide us with some insights as to the positive relationship between planning and lexical retrieval and access (Ortega, 1999). For instance, the availability of planning sheets (with the target FSs and/or the facilitative questions for essay content development) during the planning stage would greatly affect the written performance as seen in the CLFP, LFP, CFP and MG groups.

5.4 The interrelationship among all CAF measures and between the CAF measures and FS use

Results from the MANOVA suggested interrelations among the CAF measures, and the correlations matrix further confirmed the intricate relationships. One of the most interesting findings was the correlation between complexity and accuracy measures. It seemed that both the LFP and CLFP conditions led to greater complexity without sacrificing accuracy, indicating the learners did not produce language of lower accuracy when they produced language of greater syntactic and phrasal complexity. The positive relationship was consistent with the findings from the MANOVA that simultaneous

effects of planning on complexity and accuracy were observed in the CLFP group. Regarding the correlations between fluency, accuracy and complexity, Skehan and Foster (1999) propose that fluency is more related to learners' control over and automaticity of their linguistic L2 knowledge, while complexity and accuracy are primarily related to L2 knowledge representation. Not surprisingly, the low or non-significant relationships between complexity/accuracy measures and measures of fluency were similar to those observed in literature.

Furthermore, no trade-off effects were observed for any planning condition. To account for the parallel performance of accuracy and complexity, Skehan (2009) postulated that given that positive correlations between accuracy and complexity were found to be less common, it should not be the result of task complexity predicted by Robinson's Cognition Hypothesis, but rather the reflection of "the joint operation of separate task and task condition factors" (p. 510). The assumption is that participants were provided not only with target FSs in accurate forms (as presented on their planning sheets), but at the same time, the target forms were adequately complex and salient enough to "force" participants to draw on them, even though the focus of the guided planning (as in the CLFP group) was not only the language itself. An equally plausible explanation was that attentional limitations were eased when the target FSs and facilitative questions for essay content development were available on the planning sheets. Especially in the case of CLFP condition, the target FSs presented were likely to lead to a

greater accuracy, while the existence of content-centered questions contributed to the increase in greater syntactic complexity.

Actually, Foster and Skehan (1999) argued that even Robinson (1996) acknowledged that the integration of accuracy and complexity derives from the "stimulus to the performance [for instance, pre-task guidance and planning manipulation in the current study] rather than the task demands themselves" (p. 241). In other words, it was the input the participants received that prompted them to consider both accuracy and complexity, rather than how they responded to task instructions or teacher's planning manipulation. All in all, one of the major findings in the current study is to support that the joint operation observed between complexity and accuracy was more likely the result of planning effects due to task instruction, rather than task complexity itself.

Secondly, results indicated significant correlations between FS uptake, target and non-target FS combined use, and the complexity measure (clauses per TTCU). The results were contradictory to the hypothesis that participants' use of the target FSs should be positively related to total number of words (fluency). The hypothesized correlation between FS use and fluency is based on the understanding that facilitating the access and retrieval of the target FSs in the L2 mental lexicon will help learners access their exemplar-based system faster and consequently, contribute to language fluency, at least when fluency is measured in terms of number of words. In first place, the lack of relationship between fluency and FS use and uptake could be attributed the length of the

timed writing sessions. In both sessions, participants were given a total of 40 minutes to write which could be too long to observe the planning effects on fluency.

It turned out that the facilitative role of planning (with a focus on language, content or both) seemed to be the most obvious in promoting complexity. Without further data from stimulated recall which may allow for reexamining of the planning processes, the most likely explanation for the results is that learners' use of target FSs were prompted by the planning sheets which guided subsequent written production involving both syntactic complexity and accuracy. In other words, learners' use of target FSs in the written task didn't suggest their automatic control over the FS structures, but were an indicator of their analytical use of the FSs, which in turn, led to more syntactically complex and accurate language. The nature of formulaic language may make it a driving force for syntax. This is remarkable for Chinese in which the positioning of multi-word collocations, phrasal frames and sentential connectives is more flexible than that of its English counterparts. For instance, sentential connectives in Chinese can be placed in either clause-initial, predicate-initial, or clause-final positions, thus FS use in Chinese is more likely to accompany a more complex syntactic analysis.

It seems that the syntactic analysis contributed to both FS use and complex (and maybe accurate) language production, but not necessarily language fluency. In fact, the analytic use of FSs is the central tenet of Wray's (2002) model of L2 formulaic language use and acquisition. She proposed that at the time of the initial exposure to the sequences, L2 learners take apart formulaic sequences to get the lexical constituents, store them

separately, and do not keep the grammatical information (morphology in particular) about how word strings stay together. This might lead to partial recall or even errors in FS use because learners have to compose a sequence out of individual words. Wray's hypothesis was especially useful for interpreting the use of sentence frames and phrasal collocations in the study. According to Wray, it is only with more frequent exposure to the target language environment that L2 learners' FS access and retrieval would become more automatized and less controlled.

5.5 Qualitative results

Evidence of longer TTCUs (measured as mean length TTCU), to some extent, was observed in the L2 Chinese written data in the study, representing syntactic features of native speakers' of Chinese. In fact, learners do not acquire and use topic chains until they have reached a very advanced stage (Li, 2004), given that zero pronouns and topic chains are the most difficult aspect of Chinese discourse for L2 learners to use (Xiao, 2004).

5.5.1 FS use as an indicator of phrasal complexity

Depending on the categories of FSs, the general argument that FS use contributed to writing complexity could be questionable. Presumably, FS retrieval and access are closely related to syntactic complexity, at least, in L2 Chinese.

It should be noted that there were no particular lexical complexity measures such as type-token ratio adopted in the study. Rather it was hypothesized that the use and accurate use of FSs (both target and non-target) might be a more fine-grained measure of lexical or phrasal complexity for L2 Chinese only if the correlations between FS use and other CAF measures could be established.

The four planning conditions posed a clear difference on the range and categories of FSs. Participants in both the LFP and CLFP groups were pushed to use the target FSs, while in contrast, the CFP and MG groups produced significantly fewer occurrences of FSs, indicating the effects of task condition. Correlations between the FS use and uptake and the two syntactic complexity measures found in the LFP and CLFP groups further confirmed the hypothesis that FSs might be a possible measure for phrasal complexity.

It could be hypothesized that lexical/phrasal complexity measures based on FSs were unidimensional. Theoretically, it is difficult to "separate the lemma retrieval and syntactic building stages" (Skehan, 2009, p. 526) since FS access and retrieval represent a prominent feature in L2 oral and written production. It can be further postulated that the availability of FSs presented in the planning sheets of the LFP and CLFP groups were likely to ease the cognitive load, accelerate access to and retrieval of more complex lexis, and possibly, trigger the need to formulate more syntactically complex language. In other words, the LFP and CLFP conditions, as permitted by the task design, influenced "the level of language complexity appropriate for a particular task" (p. 517).

Based on the results of corpus analyses, Biber and Gray (2010) examined structural complexity at two levels: phrasal and clausal. According to Biber and Gray, phrasal features (such as nominal phrases) function to compress discourse while clausal complexity (such as coordination and subordination) functions to elaborate discourse. They argued that writing relies on both phrasal and clausal complexity, and thus using additional measure for phrasal complexity (such as FS use and uptake in the current study) would provide a more accurate picture regarding written complexity performance.

Studies (e.g., Nesselhauf, 2003) also provided empirical evidence to support the importance of lexical complexity measures. In Nesselhauf's study (2003), collocation accuracy (i.e. the words in the sample that collocate accurately together) contributed 84% of the variance in holistic scores in the written samples and 89% of the variance in the spoken samples. A recent study by Verspoor, Schmid, and Xu (2012) confirmed the validity of FS use in measuring written performance. It was found that the "number of chunks" emerged as one of the six indices (the other five are sentence length, the Guiraud index, all dependent clauses combined, all errors combined, and the use of present and past tense) that would distinguish between different levels of writing proficiency with a sample of 437 texts written by Dutch learners of L2 English. According to the authors, this was the first study that examined "chunks" systematically across proficiency levels, leading to their claims that at the early stage of acquisition, there was more syntactic reorganization, and lexical variations emerged later on to distinguish proficiency levels. In a similar vein, Crossley, Salsbury and McNamara (2009) also demonstrated that

analytic judgments of collocation accuracy, lexical diversity, and word frequency were highly predictive of holistic judgments of lexical proficiency for both written and spoken samples.

Realizing the robustness of various measures in gauging cognitive processing effects on L2 production, Robinson, Cadierno, and Shirai (2009) argued for a higher sensitivity of the specific measures based on selected dimensions of task complexity. Similarly, Michel, Kuiken and Vedder (2012) called for researchers to include task specific measures rather than global CAF measures when discussing significance related to the Cognition Hypothesis. In fact, both Skehan (2009) and Robinson (2011) agreed on the vital role of lexis measures and suggested that CAF measures need to be supplemented by measures of lexical use, even though they didn't explicitly list FSs as one subset of lexical use.

As discussed in 5.1.1., complexity is a multi-dimensional construct that serves as an indicator for L2 production and development. For this reason, it was proposed that FS based measures could be better termed as measures of phrasal complexity in order to distinguish them from lexical complexity, clausal complexity, and syntactic complexity.

5.5.2 "Fixedness" of formulaic language

Presumably, FS use is closely related to syntactic complexity in L2 Chinese as discussed above. However, the general argument that FS use contributed to writing complexity could be questionable without distinguishing FS types.

In the current study, multi-word collocations were found to be used more frequently, and more accurately than phrasal frames and sentential connectives, suggesting the structure of the FSs could have some effects on FS retrieval and access. It seemed that multi-word collocations were used more similar to single words with a higher number of occurrences and greater accuracy. However, producing phrasal frame structures and sentential connectives were more likely to trigger syntactic analysis, indicating they are not necessarily stored and processed as holistic units.

The fact that certain target FSs were not entrenched enough at the time of use is determined by the intrinsic characteristics of formulaic sequences such as length, structure, frequency, and semantic and functional saliency (see for example, Conklin & Schmitt, 2007; Schmitt, 2004). The results revealed that the most recalled (i.e. easiest) sequences in the EI test were from the category "multi-word collocation", while the least recalled, and therefore most difficult, sequences tended to be phrasal frames and sentential connectives. As discussed previously, the degree of fixedness (also referred to as "variability") along with continuity as two distinct characteristics of FS was addressed by Nattinger and DeCarrico (1992), who claimed that polywords, institutionalized expressions, phrasal constraints and sentence builders decrease in their degree of fixedness with the last category (sentence builders) being mostly non-continuous and with fillable slots. The nature of different types of FS constitutes a source of variation in language use with the open structures allowing for wider semantic possibilities, while there is understandably less variation occurring in fixed structures.

In addition, almost all multi-word collocations in Chinese are fixed structures and consist of content words, manifesting a higher degree of semantic transparency. According to Schmitt (2004), semantic transparency refers to how the meaning of the whole sequence differs from its individual parts. If the meaning of the sequence can be induced from its constituent parts, the sequence is of high semantic transparency. However, partially-fixed phrasal frames in Chinese usually consist of both content and function words with a lower degree semantic transparency. The least semantically transparent type is probably sentential connectives which are made up of purely function words.

As far as variation is concerned, few instances of restructuring of the target FSs were seen in the data, which in contrast, is a common phenomenon observed in L2 English data. It seems that Wray's (2002) model of lexical representation of classroom-taught L2 English learners might not fully account for the use and uptake of FSs in L2 Chinese, at least not for the use of phrasal collocations and sentence frames in the study. Restructuring of FSs has been interpreted by Wray as a result of "having too much choice over linguistic forms for a specific meaning (2000, p. 206)," but not necessarily the original correct sequences in their mental lexicon, so learners "reconstruct" new sequences (often with errors) based on words that made sense.

Instead, the "all-or-nothing" principle (Schmitt, 1994) works better to explain the variations of FS use in Chinese. It is acknowledged that formulaic language use is idiosyncratic to language learners and is subject to individual experience, language

exposure, language proficiency, memory capacity and the environment of language use, such as tasks (Wray & Perkins, 2000). Qualitative analysis revealed that errors of FS use in L2 Chinese were mostly tied to syntactic inappropriateness (such as verb transitivity errors, FS induced incorrect word order and missing part(s) in sentential connectives), which corresponded with Zhan's (2012) and Wang's (2013) findings that errors concerned the immediate constituents succeeding or following the FSs.

CHAPTER 6 SUMMARY AND CONCLUSION

6.1 Introduction

To date, the role and effects of planning on formulaic language use and uptake, and written performance have not been adequately discussed in the literature. The dissertation addressed a few controversies in the related theoretical frameworks and proposed the term “a lexical approach to writing instruction” with the attempt to connect the three frameworks in SLA research: FS use and uptake, task planning, and writing complexity, accuracy and fluency (CAF) measures.

Pedagogically, it is believed that with a better understanding of which planning conditions and which tasks may bring about sustained balanced development in complexity, accuracy and fluency, teachers can support learners through more rigorous planning conditions and appropriate writing tasks.

6.2 The nature of task planning

By comparing different task planning conditions, the primary contribution of the study is to examine the trade-off hypothesis supported by Skehan’s Limited Attentional Capacity model. Drawing on findings in the study, it seemed that simultaneous influences of task on complexity, accuracy and fluency can be attributed to planning manipulation and/or teacher’s task input, rather than the tasks themselves. In other words, acknowledging attentional limitations does not prevent generating predictions that complexity, accuracy and fluency would co-vary, with the understanding that complexity

and accuracy may actually work in different mechanisms due to varying task environments and constraints.

Various planning conditions examined in the study provide a valid explanation for the simultaneous influences on complexity and accuracy, thus the arguments around "limited attentional capacity" seemed trivial as we are not addressing two competing constructs in a limited capacity, but two parallel capacities at different levels due to an external factor such as task manipulation. As such, the results concerning the intricate relationships between complexity, accuracy, and fluency suggest that the CAF measures do not operate in complete independence from each other, and that findings obtained by CAF measures depend crucially on the context (e.g., task instruction/condition) in which the data are collected.

6.3 Planning as a pedagogical tool

It is undeniable that formulaic language acquisition requires learners to notice, restructure and reproduce the target forms, which can be facilitated by various conscious-raising activities, including pedagogical planning which involves providing conditions for noticing and sustained emphasis on re-noticing (e.g., Doughty & Williams, 1998; Samuda, 2001; Schmidt, 1990, 1995).

The focus of the study is whether directing learners' attention to linguistic forms (FSs in the study) in the planning stage before writing will promote the use and uptake of such vocabulary knowledge, and consequently, affect the quality of the written products

in terms of complexity, accuracy and fluency. For oral production, Mochizuki and Ortega (2008) and Sangarun (2005) reported the effectiveness of "manipulating" the participants' attention in certain ways during pre-task planning. Based on the literature discussed, it was hypothesized that guided planning before writing would make the target linguistic resources more accessible, and likely ease the cognitive load for learners during the writing processes.

The findings of this study provide evidence for the value of pre-task planning and effect of task condition in form-focused instruction (Doughty & Williams, 1998), especially for formulaic language use through written production tasks. The manipulation of task conditions directed learners' attention and triggered learner-driven focus on form, which represents a similar notion to Swain's pushed output (1985, 1995.

Overall, it supported the important role of pedagogical tasks engaging FS use in writing as much as frequently exposing students to FSs through reading and listening materials. Based on Hulstijn and Laufer's involvement load hypothesis (2001), Boers and his colleagues (Boers et al., 2006) experimented with an instructional method that emphasized 'noticing' by directing learners' attention to the "syntagmatic dimension" (p. 249) of the L2 formulaic sequences. In a following study, Boers and Lindstromberg (2012) reviewed previous studies on instructional approach towards formulaic language acquisition and reaffirmed that drawing learners' attention to formulaic sequences was not enough, and acquisition of formulaic language required repeated exposure as well as explicit learning.

Furthermore, it is postulated whether learners' allocation of attention is trainable, so in the long run, they will feel more comfortable in prioritizing their attention to linguistic forms or meaning depending on the nature of tasks. If Skehan (2009) was right concluding that the parallel increase of accuracy and complexity was the result of task input/instruction and teacher-initiated planning manipulation rather than task complexity, then the benefits of guided planning as a pedagogical intervention are easily applicable to a classroom setting.

Fuenten (2006, p. 266) agreed on the conclusions in the Ellis' (2003) study that "a more explicit, teacher generated, planned focus on form" may be more effective in directing learners' attention, especially the "structure based production task" where learners needed to use specific L2 forms in order to carry out the task. According to Fuenten, "learner-initiated focus on form was scarcely observed" (p. 286) if the target words easily went unnoticed and were ignored by learners if they were not essential for task completion. In contrast, the structure based production tasks will "allow for on-line retrieval of target words, and deeper processing of the L2 words by helping learners to establish more productive meaning–form connections through multiple opportunities for output production (of target words)" (p. 282).

One thing to be noted is whether the effectiveness of task-based focus on form could be equally applied to the acquisition of new linguistic items. Or does task-based focus on form only facilitate use of the already learned linguistic forms? This was the argument put forward by Bygate (1999) who was in favor of the facilitative role of form-

focused instruction. In the current study, all target FSs were previously taught (even though participants identified a few items as "unknown" in the pre-test VKS), thus generalization to new linguistic items needs to be made cautiously.

Furthermore, data collected from the planning sheets indicated that participants across planning groups were not able to fully utilize the opportunities of planning regardless of planning type, or learners may simply not have known how to take advantage of the planning opportunity (Elder & Iwashita, 2005). Thus, it is necessary to incorporate pedagogical interventions into the planning stage. In other words, carefully designed guidance with a specific focus on language forms or content should be made explicit by the teachers. As Samuda (2001) proposed, the ideal form-focused task will create "a semantic space" (p. 122) which pushes learners to go beyond their interlanguage repertoire and recognize "gaps" in order to be engaged meaningfully with the task demands. Thus teachers should aim at designing and implementing writing tasks which encourage learners to focus on linguistic forms and language use in the planning stage to maximize the effects of planning in L2 classrooms.

Results from the study would also inform language educators on how the effects of planning were constrained by other factors such as the nature of the linguistic forms ("fixedness" of the FSs in the present study). Formulaic sequences have been argued to be one of the strongest discriminators between levels of L2 English proficiency (Laufer & Waldman, 2011; Verspoor, Schmid, & Xu, 2012). Nevertheless, it is yet to be empirically tested whether directing learners' attention to the target FSs in the planning

stage will lead to variances in terms of FS use in written reproduction. It seemed that certain types of formulaic language, multiword structures, but not phrasal collocations and sentence frames, were used more frequently and accurately, suggesting the influence of structure of the FSs. As such, the merits of the current study lie in the fact that specific measures of syntactic complexity (clauses per TTCU) and lexical complexity (number of FSs and FS uptake) were adopted in the analysis, alongside other general measures (mean length TTCU for complexity and mean length of EFTTCUs for accuracy). It's believed that the fine-grained measures allowed for a more accurate representation of the controversial CAF relationships.

Finally, an important insight gained from the study is that teachers have options when designing writing tasks. It is commonly assumed by language teachers that if the nature of tasks relate more to communication of information than to practice of grammatical structures, learners are unlikely to direct their attention to language than to content. Nevertheless, learners' performance in the CLFP group in the current study challenged this view. The findings indicated that the CLFP condition contributed to positive results in clausal complexity, phrasal complexity (indicated by FS use and uptake), accuracy and fluency. In other words, language-focused pre-task planning, for example, is not the only option for teachers who want to design tasks that would engage students in the use and uptake of formulaic language. Other planning conditions, such as language- and content-focused planning (CLFP), if designed appropriately, are likely to be equally effective. One of the pedagogical implications of the study is that the teacher

could and should intervene in form-focused instruction and help learners to achieve an optimal balance in how they divide their attention between language and content.

However, it has to be kept in mind that the effectiveness of planning conditions varies according to learner types, and teaching outcomes depend mostly on the motivating aspects that stem from using a variety of activities. Nevertheless, the current study's results suggest that language teachers may resort to writing tasks that incorporate newly taught words in order to enhance students' uptake.

6.4 The merits of classroom-based study

This study employed a quasi-experimental design, and the writing task echoed the writing context in a real classroom setting. The fact that the effectiveness of planning was tested in a classroom setting contributed to the needed connections between research, methodology and classroom.

Firstly, given that there is a lack of detailed information and consistency regarding how FSs were scored and coded in previous studies (e.g., Boers et al., 2006; Foster, 2001; Rott, 2009) around the effectiveness of planning as awareness-raising on FS use, the current study proposed a detailed analysis of FS categorization and coding, which could benefit teachers in examining L2 Chinese writings involving FSs use in the future. In addition, the preparation and use of planning sheets and prompts were addressed, highlighting the needs for carefully constructed writing instructions by teachers in designing form-focused planning activities before writing.

Furthermore, neither Boers et al. (2006) nor Rott (2009) reported controlling time as a variable in their studies on the role of awareness-raising activities on formulaic language acquisition. The current study tested the hypothesis that in class timed-writing could be used as a pedagogical intervention to facilitate written output performance (on complexity measures, at least). The conclusion added to the understanding that when designing form-focused planning activities, time needs to be considered as a contextual variable.

Lastly, the robustness of TTCU based measures in the current study indicated that T-unit based measures which have been employed in previous studies of L2 English might not be valid in measuring Chinese written production for advanced L2 Chinese writers. Taken together, teachers should consider employing T-unit based measures (for beginners and intermediate learners) and TTCU based measures (for the more advanced learners) in analyzing L2 Chinese written texts based on objective measures.

6.5 Limitations

With regard to interlanguage development and CAF measures, Larsen-Freeman called for more longitudinal and non-linear research. In a similar vein, Norris and Ortega (2009) doubted whether a linear or co-linear trajectory of accuracy, fluency, and complexity would represent L2 acquisition or development, and argued for multi-dimensionality, dynamicity, variability, and non-linearity in future CAF research.

One objective of this study was to use a repeated-measures approach to determine if L2 writers' use of FSs in writing would change according to planning condition. To reduce the potential effect of having them write on the same topic in a different condition, a second topic was selected that was considered comparable in nature and equally approachable for the writers. However, the fact that the current study is not a longitudinal one wouldn't guarantee a dynamic and accurate observation on FS use and uptake.

Regarding the explicit role of planning in directing attention to form or content during the pre-writing stage, even though the participants' planning sheets were collected, few conclusions could be drawn, unfortunately, due to the fact that most of the planning sheets were left blank. The absence of think-aloud, stimulated recall, or any retrospective data constrained further hypotheses on the cognitive and attentional processes engaged in during planning. As noted by Slimani (1991), learners' focus of attention can be rather different from the focus intended by teachers, and that learners can engage in tasks and interpret them in ways that differ from those predicted or desired (e.g., Coughlan & Duff, 1994; Kumaravadivelu, 1993; LaPierre, 1994). Thus evidence of learners' actual use of planning opportunities and allocation of attentional resources needs to be better documented.

Finally, unlike data found for L2 English learners, few instances of FS restructuring were observed in the L2 Chinese data, indicating a limited range of variation. The possible "all-or-nothing" principle to account for less restructuring needs to be further investigated with a future study that is more qualitative in nature.

Nevertheless, findings in the study suggested that learners from the CLFP group were likely to attend to both form and meaning in guided pre-writing planning. As such, the most significant contribution of the dissertation is that it provides additional evidence regarding mixed results over trade-offs between CAF in the literature. The results support Skehan's recent modification on the Limited Attentional Capacity model with further distinctions between task instruction/condition and task complexity regarding L2 written language production.

What's more, the hypothesis has been tested that the learning of L2 Chinese formulaic language could be facilitated through writing instruction with carefully designed writing prompts and pre-writing guidance. It is likely that planning with a focus on form will prompt learners to use more target FSs and use them more accurately which echoes L2 learners' variation in complexity, accuracy and fluency.

APPENDICES

Appendix I *Demographic Questionnaire*

Name _____ Age _____

Gender (circle one): Female / Male

Native language(s):_____

1. How long have you been learning Chinese? _____ month(s)

2. Where did you learn Chinese before you came to China? And how long?

   _____ month(s) in high school

   _____ month(s) in college/university

3. How long have you been in China? _____ month(s)

4. Have you ever studied Chinese in a Chinese speaking country before this semester?

   _____ Yes _____ No

   If yes, please specify where you have studied in China _____

   And how long? _____ month(s)

5. If you watch TV in Chinese, how many hour per week? _____

6. If you listen to music in Chinese, how many hour per week? _____

7. If you read (magazines, newspapers, novels and websites etc.) in Chinese, how many hour per week? _____

8. How often will you interact or have conversations with a native Chinese speaker?

   _____

Appendix Table 1 *The Vocabulary Knowledge Scale* (participants' version in Chinese)

下面这些短语都是课文中的。请在 A、B、C 、D 中选择符合最自己的情况的一项，并且在短语旁边的方框中划勾或者造句。（注意：请不要使用字典或电子字典）

	A 我不知道这个短语	B 我知道这个短语，可是不会用	C 我知道这个短语，而且可以翻译	D 我会用这个短语造句_____ (请你写下一个句子。可以写在这张字的背面)
家喻户晓				
独具特色				
世井文化				
侃侃而谈				
温文尔雅				
精明能干				
熙熙攘攘				
鱼龙混杂				
无伤大雅				
有点儿				
满不在乎				
由它去				
闲得无聊				
有说有笑				
凑热闹				
自由自在				
悠然遐想				
对……而言				
有…….之称				
类似于				
迁怒于				
…居多				
说不上…				
哪怕 …也				
不仅...也				
无论…都				
…也好…也好				
是…还是…				
若…就…				
一…就…				

The translated VKS Appendix Table 1 (cont'd)

	I don't know the phrase	I have seen the phrase before but I'm not quite sure about the meaning	I understand this phrase and I can translate the phrase.	I can use this phrase to make a sentence_____ (Please write down your sentence).
家喻户晓 a household name				
独具特色 unique and peculiar				
世井文化 urban bourgeois culture				
侃侃而谈 to talk non-stop				
温文尔雅 refine and cultivated				
精明能干 astute and shrewd				
熙熙攘攘 hustle and bustle				
鱼龙混杂 mixture of (good and bad)				
无伤大雅 without bothering sb/sth				
有点儿 a little bit				
满不在乎 do not care at all				
由它去 let it be				
闲得无聊 feel bored being idle				
有说有笑 talking and laughing				

The translated VKS Appendix Table 1 (cont'd)

凑热闹 come along for the fun				
自由自在 carefree and unstrained				
悠然遐想 to think leisurely and be lost in reverie				
对……而言 means…to sb/sth				
有…….之称 has the title /name of				
类似于 be similar to				
迁怒于 vent one's anger on sb/sth				
…居多 the majority is…				
说不上… cannot tell whether…				
哪怕 …也 even if…still				
不仅...也 not only…, but…				
无论…都 no matter…still…				
…也好…也好 no matter (this) …and (that)…				
是…还是… no matte…or…				
若…就… if…then…				
一…就… once…then				

Appendix II *Topic 1 and writing prompt*

Instructions: You will have one minute to read the task prompt. After that you will have 10 minutes to read the planning sheet and to plan according to what you have read for your writing. You may note down your plan in English or Chinese, but do not write out everything in detail. You will not be allowed to use your notes on the planning sheets when the planning time is over. Please put your planning notes away and start to write immediately.

Topic: A special city/place

Audience: Anyone who will be interested to know more about a city/place that has a special meaning to you.

Purpose: A reflective essay helps you reflect critically on something that you've read, learned, observed, felt, or experienced. It is possible that you draw some conclusions about the significance of those experiences in relation to the context of your lives. So your purpose is to let your readers know about what your feel and think about the topic, where those feelings and thoughts come from and how they shape your views.

Time: You will have 30-minute class time to finish the writing.

Length: Ideally your writing should be about 400 characters which allow you to provide a full examination and summary of what you think about the topic.

Evaluation: Your essay will be evaluated in terms of comprehensiveness of the content, length of the essay and grammatical accuracy.

Appendix III *Topic 2 and writing prompt*

Instructions: You will have one minute to read the task prompt. After that you will have 10 minutes to read the planning sheet and to plan according to what you have read for your writing. You may note down your plan in English or Chinese, but do not write out everything in detail. You will not be allowed to use your notes on the planning sheets when the planning time is over. Please put your planning notes away and start to write immediately.

Topic: Your lifestyle or habits

Audience: Anyone who will be interested to know more about your lifestyle or habits.

Purpose: A reflective essay helps you reflect critically on something that you've read, learned, observed, felt, or experienced. It is possible that you draw some conclusions about the significance of those experiences in relation to the context of your lives. So your purpose is to let your readers know about what your feel and think about the topic, where those feelings and thoughts come from and how they shape your views.

Time: You will have 30-minute class time to finish the writing.

Length: Ideally your writing should be about 400 characters which allow you to provide a full examination and summary of what you think about the topic.

Evaluation: Your essay will be evaluated in terms of comprehensiveness of the content, length of the essay and grammatical accuracy.

Appendix IV *Topic 1 planning sheets*

Pre-writing planning sheets (CFP) Topic 1
Directions: The following questions are from the reading texts we have been engaged with this week. Preparing to answer these questions may help you plan the writing more easily. You are free to rearrange the order of the questions in order to develop body paragraphs that flow together for the reader.
(Questions with the target phrases embedded)

1. “总有一个地方，会在无意中保留你童年的乐园”。你童年的乐园在什么地方？这个地方对你而言意味着什么？哪怕你已经离开很久，你也会怀念你的乐园吗？
There will always be a place as your childhood dreamland. Where it is for you? What does this place mean to you? Even if you’re far from the place, will you miss it?

2. 你想要介绍的这个特殊的地方，是否有很长时间的历史? 是否家喻户晓？是否独具特色？
Does the place you’re going to introduce have a long history? Is it a household name? How is it unique?

3.城隍庙是上海市井文化的化身，课文当中有 “上海的襁褓”之称。你想要介绍的地方代表了一种什么样的文化？这种文化类似于上海的市井文化吗？
Chenghuangmiao is the symbol of urban bourgeois culture and enjoys the title of “the birthplace of Shanghai”. What kind of culture does the place you’re introducing represent? Is it similar to the Shanghai urban culture?

4.北京人喜欢侃侃而谈。和北京人相比，上海人显得更加温文尔雅、精明能干。你想要介绍的这个特殊的地方，人们在性格、语言、生活习惯等方面有什么特点？
Native Beijing residents are quite talkative. Compared to them, Shanghai locals are more refined and shrewd. What are the characteristics of the people (e.g., their language and life style) of the place you’re introducing?

5.有的人喜欢大城市熙熙攘攘、鱼龙混杂的生活，而有的人更喜欢农村的生活。还有的人不仅喜欢城市也喜欢农村，无论哪种生活都喜欢。城市也好，农村也好，对你而言，你更喜欢在哪儿生活？
A lot of people enjoy the hustle and bustle and mixed life styles of urban cities, while some prefer to live in the countryside. Others like not only the urban but also the country lives. They enjoy it no matter what life style it is. Which one do you prefer?

Pre-writing planning sheets (LFP) Topic 1
Directions: The following vocabulary list is comprised of four-word phrases, multi-word structures, idioms and sentence frames from this week's reading texts. These phrases may help you connect your information so that it is easier for your readers to follow. Try to join these phrases to the writing of the reflective essay when you are preparing for the planning.

1. 对……而言 [means…to sb/sth]　　哪怕 …也 [even if…]

2. 家喻户晓 [a household name]　　独具特色 unique and peculiar]

3. 世井文化[urban bourgeois culture]　有…….之称 has the title /name of]
   类似于 [be similar to]

4. 侃侃而谈 [to talk non-stop]　　温文尔雅 [refine and cultivated]
   精明能干 [astute and shrewd]

5. 熙熙攘攘　[hustle and bustle]　　鱼龙混杂　[mixture of (good and bad)]
   不仅...也 [not only…, but…]　　无论…都[no matter…still…]
   也好…也好　[no matter (this) …and (that)…]

Pre-writing planning sheets (CLFP) Topic 1
Directions: The following questions and vocabulary are from the reading texts we have been engaged with this week. Preparing to answer these questions may help you plan the writing more easily. The vocabulary may help you connect your information so that it is easier for your readers to follow.

1.总有一个地方，会在无意中保留你童年的乐园”。你童年的乐园在什么地方？
There will always be a place as your childhood dreamland. Where it is for you?

2. 你想要介绍的这个特殊的地方，是否有很长时间的历史？是否家喻户晓？是否独具特色？
Does the place you're going to introduce have a long history? Is it a household name? How is it unique?

3. 城隍庙是上海市井文化的化身。你想要介绍的地方代表了一种什么样的文化？
Chenghuangmiao is the symbol of urban bourgeois culture of Shanghai. What kind of culture does the place you're introducing represent?

4. 你想要介绍的这个特殊的地方，人们在性格、语言、生活习惯等方面有什么特点？
What are the characteristics of the people (e.g., their language and life style) of the place you're introducing?

5. 有的人喜欢大城市的生活，而有的人更喜欢农村安静淡薄的生活。对你而言，你更喜欢在哪儿生活？
A lot of people enjoy living in the big cities, while some prefer to live in the countryside. Which one do you prefer?

~~~~~~

1 对……而言 [means…to sb/sth]　　哪怕 …也 [even if…]

2 家喻户晓 [a household name]　　独具特色 unique and peculiar]

3 世井文化[urban bourgeois culture]有…….之称 has the title /name of]　　类似于 [be similar to]

4 侃侃而谈 [to talk non-stop]　　温文尔雅 [refine and cultivated]　　精明能干 [astute and shrewd]

5 熙熙攘攘 [hustle and bustle]　　鱼龙混杂 [mixture of (good and bad)]
不仅...也 [not only…, but…]　　无论…都[no matter…still…]
也好…也好 [no matter (this) …and (that)…]
~~~~~~

Appendix V *Topic 2 planning sheets*

Pre-writing planning sheets (CFP) Topic 2
Directions: The following questions are from the reading texts we have been engaged with this week. Preparing to answer these questions may help you plan the writing more easily. You are free to rearrange the order of the questions in order to develop body paragraphs that flow together for the reader.
(Questions with the target phrases embedded)

1.很多人认为抽烟是见无伤大雅的习惯。你对抽烟的人有点儿反感吗？还是满不在乎，由它去？
Some people think smoking in the public has nothing to do with others. Are you a little bit annoyed by this attitude? Or you do not care about it at all and just let it be?

2.课文中把抽烟比喻为一个“伴儿”。无论你是闲得无聊，还是跟朋友有说有笑，都可以抽烟。你介绍的这种习惯也是你的“伴儿”吗？若生气了便可以迁怒于它吗？
Smoking has been mentioned as a “partner” in the text. No matter whether you’re bored or enjoying the company of friends, you can always smoke. Is the habit you’re going to introduce like your partner? If you’re anger, will you vent your anger through smoking?

3.你要介绍的这种生活习惯是在朋友中流行的吗？这些朋友是什么样的人？单身的朋友居多吗？喜欢凑热闹、说闲话吗？
Is the habit you’re introducing popular among your friends? What are those people? Is the majority of your friends single? Do you enjoy come along just for fun?

4. 有人喜欢生活得自由自在，这样的生活说不上好与不好。你的朋友和家人赏识你的生活方式和生活习惯吗？
Some people live in a carefree and unstrained way. It’s hard to judge whether it is good or bad. Do your friends and family member appreciate your habits and life style?

5.假设你正在悠然遐想，你可以选择一种新的生活习惯，你会选择什么？如果你一做选择，就不可改变呢？
Imagine you’re now leisurely immersed in your own thoughts that you’re about to change your life habits. What will you want to change? What if once you make your choice, it cannot be retrieved?

Pre-writing planning sheets (LFP) Topic 2

Directions: The following vocabulary list is comprised of four-word phrases, multi-word structures, idioms and sentence frames from this week's reading texts. These phrases may help you connect your information so that it is easier for your readers to follow. Try to join these phrases to the writing of the reflective essay when you are preparing for the planning.

1. 无伤大雅[without bothering sb/sth] 派头 有点儿[a little bit]
   满不在乎[do not care at all] 由它去[let it be]

2. 无论 还是[no matte…or…] 闲得无聊[feel bored being idle]
   有说有笑[talking and laughing] 若…就…[if…then…]
   迁怒于[vent one's anger on sb/sth]

3. 单身 …居多[is the majority of…] 凑热闹[come along for the fun]
   说闲话 像…一般

4. 自由自在[carefree and unstrained] 捉摸不定 说不上[…cannot be told]
   赏识 不妨

5. 除非 必然 悠然遐想[be leisurely immersed in one's own thoughts]
   可喜的是 一…就… [once…then…]

Pre-writing planning sheets (CLFP) Topic 2

Directions: The following questions and vocabulary are from the reading texts we have been engaged with this week. Preparing to answer these questions may help you plan the writing more easily. The vocabulary may help you connect your information so that it is easier for your readers to follow.

1.你对抽烟反感的人吗？
Are you a little bit annoyed by smoking in the public?

2. 课文中把抽烟比喻为一个“伴儿”。你介绍的这种习惯也是你的“伴儿”吗？
Smoking has been mentioned as a “partner” in the text. Is the habit you’re going to introduce like your partner?

3. 你要介绍的这种生活习惯是在朋友中流行的吗？这些朋友是什么样的人？他们的生活是怎样的？
Is the habit you’re introducing popular among your friends? What are those people? What are their life styles?

4. 你的朋友和家人赏识你的生活方式和生活习惯吗？
Do your friends and family member appreciate your habits and life style?

5. 如果可以选择一种新的生活习惯，你会选择做什么？如果你一做选择，就不可改变呢？
Now you’re about to change your life habits. What will you want to change? What if once you make your choice, it cannot be retrieved?

~~~~~~

1. 无伤大雅 [without bothering sb/sth]　　有点儿 [a little bit]
   满不在乎 [do not care at all]　　由它去 [let it be]
2. 无论…还是…　[no matte…or…]　　闲得无聊 [feel bored being idle]　有说有笑 [talking and laughing　　若…便…[if…then…]
   迁怒于 [vent one’s anger on sb]
3. …居多 [is the majority of…]　　凑热闹 [come along for the fun]
4. 自由自在 [carefree and unstrained]　　说不上 …[…cannot be told]
5. 悠然遐想 [be leisurely immersed in one’s own thoughts]
   一…就… [once…then…]
~~~~~~

Appendix VI *Error Classification Scheme*

a. Only code errors that are syntactic. Do not count character errors.

b. If sentence at the end of an essay is not finished, don't code it.

c. Code errors so that sentence is changed minimally. If there are two possible errors requiring equal change, code the first error.

d. Be conservative about counting comma errors; don't count missing commas between clauses or after prepositional phrases.

e. Sometimes a T-unit or TTCU can be grammatically correct on its own, but becomes odd or incoherent in a sentence. For coding purposes, do not count the units with grammatically correct zero anaphora or topic chains as errors, but count the sentence error when lacking coherence or semantic inappropriateness.

REFERENCES

REFERENCES

ACTFL Proficiency Guidelines. (2012). Retrieved from http://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012/chinese/simplified-characters/%E5%86%99%E4%BD%9C

Becker, J. (1975). The phrasal lexicon. *Bolt Beranek and Newman Report* No. 3081, AI Report No. 28.

Bardovi-Harlig, K., & Bofman, T. (1989). Attainment of syntactic and morphological accuracy by advanced language learners. *Studies in Second Language Acquisition, 11*, 17–34.

Becker, A. (2006) A review of writing model research based on cognitive processes. In A. Horning & A. Becker (Eds.), *Revision: History, Theory and Practice*. Parlor Press: Anderson, SC.

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9,* 2-20.

Bitchener, J., & Ferris, D. (2012). *Written corrective feedback in second language acquisition and writing.* New York: Routledge.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research 3*, 245-261.

Boers, F., & Lindstromberg, S. (2012.) Experimental and intervention studies on formulaic sequences in a second language. *Annual Review of Applied Linguistics, 32*, 83-110.

Bolander, M. (1989). Prefabs, patterns and rues in interaction? Formulaic speech in adult learners' L2 Swedish. In K. Hyltenstam & L.K. Obler (Eds.) *Bilingualism Across* the Lifespan (pp. 73-86). Cambridge: Cambridge University Press.

Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum,1*, 1-14.

Bygate, M. (1999). Quality of language and purpose of task: patterns of learners' language on two oral communication tasks. *Language Teaching Research, 3*, 185–214.

Bygate, M., Skehan, P. & Swain, M. (Eds). (2001). *Researching pedagogical tasks: second language learning, teaching, and assessment.* London: Pearson.

Conklin, K., & Schmitt, N. (2007). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics, 28,* 1-18.

Coughlan, P., & Duff, P. (1994). Same task, different activities: Analysis of a SLA task from an Activity Theory perspective. In J. Lantolf & G. Appel (Eds.), *Vygotskian perspectives on second language research* (pp. 173-193). New Jersey: Ablex.

Crookes, G. (1989).Planning and interlanguage variation. *Studies in Second Language Acquisition,11,* 367-383.

Crossley, S. A, Salsbury, T., & McNamara, D. S. (2009). Measuring second language lexical growth using hypernymic relationships. *Language Learning, 59*(2), 307-334.

De Larios, J., Marin, J., & Murphy, L. (2001). A temporal analysis of formulation processes in L1 and L2 writing. *Language Learning, 51,* 497-538.

Doughty, C., & Williams, J. (Eds.). (1998). *Focus on form in classroom second language acquisition.* New York: Cambridge University Press.

Elder, C., & Iwashita, N. (2005). Planning for test performance: Does it make a difference? In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 219−238). Philadelphia: John Benjamins.

Ellis. R. (2001). Non-reciprocal tasks, comprehension and second language acquisition. In M. Bygate, P. Skehan, & M. Swain, (Eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 49-74). London: Longman.

Ellis, R. (2003). *Task-based Language Learning and Teaching.* Oxford: Oxford University Press.

Ellis, R. (2005). *Planning and Task Performance in a Second Language*. John Benjamins.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics, 30*, 474-509.

Ellis, R., & Yuan, F. (2004). The effects of planning on fluency, complexity and accuracy in second language narrative writing. *Studies in Second Language Acquisition, 26*, 59-84.

Fei, F. (2011, March). *The role of cognitive factors in the representation of formulaic language in L2 mental lexicon.* Poster session at the Annual Conference of American Association of Applied Linguistics, Chicago, IL.

Field, A. (2005). *Discovering Statistics Using SPSS.* Sage: Thousand Oaks.

Foster, P., & Skehan, P. (1996). The influence of planning on performance in task-based learning. *Studies in Second Language Acquisition, 3*, 299–324.

Foster, P. & Skehan, P. (1999).The influence of source of planning and focus of planning on task-based performance. *Language Teaching Research, 3*, 215–247.

Foster, P. (2000). *Attending to message and medium: The effects of planning time on the task-based language performance of native and non-native speakers*. Unpublished Doctoral Thesis, King's College, London.

Foster, P. (2001). Rules and routines: a consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan & M. Swain (Eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 75–93), Longman: Harlow.

Foster, P., & Tavakoli, P. (2009). Native speakers and task performance: Comparing effects on complexity, fluency, and lexical diversity. *Language Learning, 59*(4), 866-896.

Fan, Y., & Bai, Y. (2007). 中高级阶段留学生汉语写作偏误分析 [Error analysis in L2 Chinese writing among intermediate and advanced learners]. *语言教学研究, 9*, 99-100.

Gao, Y. (2008). *对外汉语词汇教学 [Teaching Vocabulary in Chinese as a Foreign Language]*. Shanghai, China: East China Normal University Press.

Hopper, P. J. (1998). Emergent grammar. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 155–176). Mahwah, NJ: Erlbaum.

Housen, A., Kuiken, F. & Vedder, I. (2012). Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA. Philadelphia: John Benjamins.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161–188). Oxford: Oxford University Press.

Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in. *Language Learning, 51*, 539–558.

Hunt, K. W. (1965). *Grammatical Structures Written at Three Grade Levels*. NCTE Research Report No. 3. Champaign, IL. USA: NCTE.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. Monographs of the Society for Research in Child development, University of Chicago Press.

Ishiwaka, T. (2007). The effect of manipulating task complexity along the [+/-Here-and-Now] dimension on L2 written narrative discourse. In M. P. Garcia Mayo (Ed.), *Investigating tasks in formal language learning* (pp. 136-156). Clevedon, UK: Multilingual Matters.

Jiang, J. (2009). Designing pedagogical materials to improve awareness and productive use of L2 collocations. In A. Barfield & H. Gyllstad (Eds.), Researching Collocations in Another Language: Multiple Interpretations (pp. 99-113), Hampshire, English: Palgrave Macmillan.

Jiang, W. (2013). Measurements of development in L2 written production: the case of L2 Chinese. *Applied Linguistics, 34*(1), 1-24.

Jiang, N., & Nekrasova, T. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal, 91* (3), 433-445.

Jim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 61:* Suppl. 1, 100–140.

Jin, H. G. (1994). Topic-prominence and subject-prominence in L2 acquisition: Evidence of English-to-Chinese typological transfer. *Language Learning, 44*, 101-122.

Jin, H. G. (2007). Syntactic maturity in second language writing: A case of Chinese as a foreign language (CFL*). Journal of the Chinese Language Teachers Association, 42*, 27-55.

Johnson, M., Mercado, L., & Acevedo, A. (2012). The effect of planning sub-processes on L2 writing fluency, grammatical complexity, and lexical complexity. *Journal of Second Language Writing, 21*, 264–282

Kawauchi, C. (2005). The effects of strategic planning on the oral narratives of learners with low and high intermediate proficiency. In R. Ellis (Ed.), *Planning and Task*-Performance in a Second Language. John Benjamins.

Kim, Y. (2011). The role of task-induced involvement and learner proficiency in L2 vocabulary acquisition. *Language Learning, 61*, 100-140.

Kim, Y. (2012). 中级汉语精读课语块分析及教学初探 [Analysis on formulaic sequences in intermediate level Chinese intensive reading classes and its implications for teaching]. 山东大学硕士论文.

Kim, Y. (2013). Effects of pretask modeling on attention to form and question development. *TESOL Quarterly, 47* (1), 8-35.

Kuiken, F., & Vedder, I. (2007). Task complexity and measures of linguistic performance in L2 writing. *International Review of Applied Linguistics, 45*, 261-284.

Kuiken, F., & Vedder, I. (2008). Cognitive task complexity and written output in Italian and French as a foreign language. *Journal of Second Language Writing, 17*, 48-60.

Kumaravadivelu, B. (1993). Maximizing learning potential in the communicative classroom. *ELT Journal, 47*(1), 12-21.

LaPierre, D. (1994). Language output in a cooperative learning setting: Determining its effects on second language learning. MA thesis, Ontario Institute for Studies in Education: University of Toronto.

Laufer, B. & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English. *Language Learning, 61* (2), 647–672.

Li, C. N., & Thompson, S. A. (1976). Subject and topic: A new typology of language. In C. N. Li (Ed.), *Subject and topic* (pp. 457-490). New York: Academic Press, Inc.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar.* Berkeley: University of California Press.

Li, X. (Ed.). (2005). *Boya Chinese Intermediate Spurt I.* Beijing: Peking University Press.

Liao, J. (2010). The impact of interactive discussions on L2 Chinese composition writing. Unpublished Doctoral Dissertation. University of Iowa.

Littlewood, W. (2004). The task-based approaches: Some questions and suggestions. *ELT Journal, 58* (4), 319-326.

Levelt, W. (1989). *Speaking: From Intention to Articulation*. Cambridge, MA.

Lewis, M. (1993).*The lexical approach: The state of ELT and the way forward.* Hove, England: Language Teaching.

Lewis, M. (1997). Pedagogical implications of the lexical approach. In J. Coady & T. Huckin (Eds.), *Second Language Vocabulary Acquisition* (pp. 255-270). Cambridge, UK: Cambridge University Press.

Lewis, M. (Ed.) (2000). *Teaching Collocation: Further Developments in the Lexical Approach.* Hove, England: Language Teaching Publications.

Ma, X. (2010). 语块理论与对外汉语教学 [Theories on formulaic sequences and its implications in teaching Chinese as a second/foreign language]. 辽宁师范大学硕士论文.

McLaughlin, J., Osterhout, L., & Kim, A. (2004) Neural correlates of second-language word learning: minimal instruction produces rapid change. *Nature Neuroscience 7*(7), 703-704.

Mehnert, U. (1998). The effects of different lengths of time for planning on second language performance. *Studies in Second Language Acquisition, 20*, 83-108.

Michel, M., Kuiken, F. & Vedder, I. (2007). The influence of task complexity and interactivity. *International Review of Applied Linguistics, 45*, 241-259.

Mochizuki, N. & Ortega, L. (2008). Balancing communication and grammar in beginning level foreign language classrooms: A study of guided planning and relativization. *Language Teaching Research, 12*, 11–37.

Nassji, H. (2002). Schema theory and knowledge-based processes in second language reading comprehension: A need for alternative perspectives. *Language Learning, 52*, 439-481.

Nattinger, J. R., & DeCarrico, J.S. (1992). Lexical Phrases and Language Teaching. Oxford, UK: Oxford University Press.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics, 24*, 223–242.

Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*, 555−578.

Nunn, R. (2006). Designing holistic units for task-based learning. *Asian EFL Journal, 8* (3).

Ong, J., & Zhang, J. (2010). Effects of task complexity on the fluency and lexical complexity in EFL students' argumentative writing. *Journal of Second Language Writing, 19*, 218-233.

Ong, J., & Zhang, J. (2013). Effects of the manipulation of cognitive processes on EFL writers' text quality. *TESOL Quarterly, 47*(2), 375-398.

Oppenheim, N. (2000). The importance of recurrent sequences for nonnative speaker fluency and cognition. In H. Riggenbach (Ed.), *Perspectives on fluency*. Ann Arbor: The University of Michigan Press.

Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition, 21*, 109–48.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*, 492-518.

Ortega, L. (2005). What do learners plan? Learner-driven attention to form during pretask planning. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language*. John Benjamins.

Ortega, L. (2007). Meaningful practice in foreign language classrooms: A cognitive-interactionist perspective. In R. DeKeyser (Ed.), *Practice in a second language* (pp. 180–207). New York: Cambridge University Press.

Pallotti, G. (2009). CAF: Defining, refining and differentiating constructs. *Applied Linguistics, 30* (4), 590-601.

Park, S. (2010). The influence of pre-task instructions and pre-task planning on focus on form during Korean EFL task-based interaction. *Language Teaching Research, 14* (1), 9-26.

Pawley, A., & Syder, F.H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C Richards & R.W. Schmidt (Eds.), *Language and* Communication (pp.191-226). New York: Longman.

Polio, C. (1997). Measures of linguistic accuracy in second language writing research. *Language Learning, 47*, 101-143.

Qian, X. (2008). 汉语语块研究初探 [A preliminary exploration on formulaic sequences in Chinese].《北京大学学报(哲学社会科学版)》第 5 期.

Robinson, P. (1996). *Consciousness, Rules, and Instructed Second Language Acquisition.* New York: Peter Lang.

Robinson, P. (2001). Task complexity, cognitive resources, and syllabus design: A triadic framework for investigating task influences on SLA. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 287–318). New York: Cambridge University Press.

Robinson, P. (2003). The cognition hypothesis, task design, and adult task-based language learning. *Second Language Studies, 21*, 45–105.

Robinson, P. (2005). Cognitive complexity and task sequencing: A review of studies in a Componential Framework for second language task design. *International Review of Applied Linguistics, 43*, 1–33.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics, 45*, 193–213.

Robinson, P. (2007). Task-Based Language Learning: A Review of Issues. *Language Learning, 61 (s1),* 1–36.

Robinson, P. (2011). Task-based language learning: A review of issues. In P. Robinson (Ed.), *Task-Based Language Learning* (pp.1-36). Oxford: Wiley-Blackwell.

Robinson, P., Cadierno, T., & Shirai, Y. (2009). Mind, time, and motion: Measuring the effects of the conceptual demands of tasks on second language speech production. *Applied Linguistics, 30*, 533–554.

Robinson, P., & Gilabert, R. (2007). Task complexity, the cognition hypothesis and second language learning and performance. *International Review of Applied Linguistics,* 45*(3),* 161-176.

Rott, S. (2004). A comparison of output interventions and un-enhanced reading conditions on vocabulary acquisition and text comprehension. *The Canadian Modern Language Review, 61*(2), 169–202.

Rott, S. (2009). The effect of awareness-raising on the use of formulaic constructions. In R. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (Eds.) *Formulaic language: Acquisition, Loss, Psychological Reality and Functional Explanations* (pp. 405-423). John Benjamins.

Samuda, V. (2001.) Guiding relationships between form and meaning during task performance: the role of the teacher. In M. Bygate, P. Skehan, & M. Swain, (Eds.) *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 119-140). London: Longman.

Samuda, V., & Bygate, M. (2008). *Tasks in second language learning*. Basingstoke, UK: Palgrave Macmillan.

Sangarun, J. (2001). *The effects of Pre-task Planning on Foreign Language Performance.* Unpublished Doctoral Dissertation, OISE, University of Toronto.

Sangarun, J. (2005). The effects of focusing on meaning and form in strategic planning. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language*. John Benjamins.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics 11*, 129−158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 1−63). Honolulu: University of Hawai‘i Press.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3−32). New York: Cambridge University Press.

Schmitt, N. (2004). (Ed.). Formulaic Sequences: Acquisition, Processing and Use. Amsterdam: John Benjamins.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action: An introduction. In N. Schmitt (Ed.), Formulaic Sequences: Acquisition, processing and Use (pp. 1-22). Amsterdam: John Benjamins.

Schoonen, R., Gelderen, A., Glopper, K., Hulstijn, J., Simis, A., & Snellings, P. (2003). First Language and Second Language Writing: The Role of Linguistic Knowledge, Speed of Processing, and Metacognitive Knowledge. *Language Learning, 53*, 165-202.

Sheen, Y. (2011). The Effect of Focused Written Corrective Feedback and Language Aptitude on ESL Learners' Acquisition of Articles. *TESOL Quarterly, 41* (2), 255–283.

Shi, D. X. (1989). Topic chain as a syntactic category in Chinese. *Journal of Chinese Linguistics, 17*(2), 223-262.

Shi. D. X. (2000). Topic and topic-comment constructions in Mandarin Chinese. *Language, 76*(2), 383-408.

Sinclair, J., & Renouf, A. (1998). A lexical syllabus to language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and language teaching* (pp. 140-160). Longman, London.

Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied linguistics, 17*, 38-26.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics, 30* (4), 510–532.

Skehan, P. ,& Foster, P. (1997). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research, 1*, 185–211.

Skehan, P., & Foster, P. (1999). The influence of task structure and processing conditions on narrative retellings. *Language Learning, 49*, 93–120.

Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 183–205). New York: Cambridge University Press.

Skehan, P., & Foster, P. (2005). Strategic and on-line planning: The influence of surprise information and task time on second language performance. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 193–218). Amsterdam: John Benjamins.

Slimani, A. (1991). Evaluation of classroom interaction. In J.C. Alderson and A. Beretton (Eds.) *Evaluating language education.* Cambridge: Cambridge University Press.

Sonbul, S. & Schmitt, N. (2013). Explicit and implicit lexical knowledge: Acquisition of collocations under different input conditions. *Language Learning, 63*, 1, 121–159.

Song, W. (2009). 语块与句子加工 [Formulaic sequences and sentence processing]. 华东师范大学硕士论文.

Siyanova-Chanturia, A., Conklin, K., and Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and nonnative speakers. *Second Language Research 27*(2), 1-22.

Siyanova-Chanturia, A., Conklin, K., & Van Heuven, W. J. B. (2011). Seeing a phrase "time and again" matters: the role of phrasal frequency in the processing of multiword sequences. *Journal of experimental psychology: Learning, memory, and cognition, 37*(3), 776–784.

Su, D. (2010). 试析“构式—语块”教学法—以存现句教学实验为例 [An exploratory analysis of the "construction-chunk" pedagogy – The case of an experimental study with existential sentences]. *汉语学习 Chinese Language Learning, 2*, 83-90.

Su, D. (2011). 构式语块教学法的实质—以兼语句教学及实验为例 [The nature of the "construction-chunk" pedagogy – The case of an experimental study with Bi-constituent sentences]. *语言教学与研究 Language Teaching and Research, 2*, 16-22.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), Input in second language acquisition (pp. 235-253). Rowley: Newbury House.

Swain. M. (1995). Three functions of output in second language learning. In G. Cook, H.Widdowson & B. Seidelhofer (Eds.), Principle & practice in applied

linguistics: Studies in honor of H.G. Widdowson (pp. 125-144). Oxford: Oxford University Press,

Tavakoli, P., & Foster, P. (2008). Task design and L2 performance : The effect of narrative type on learner output. *Language Learning, 58*, 439-473.

Towell, R. (2012). Complexity, Accuracy and Fluency from the perspective of psycholinguistic Second Language Acquisition research. In Housen, A., Kuiken, F. & Vedder, I. (eds). Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA. Philadelphia: John Benjamins.

Tsao, F. (1979). *A functional study of topic in Chinese: The first step towards discourse analysis.* Taipei: Taiwan: Student Book Co., Ltd.

Tsao, F. (1990). *Sentence and clause structure in Chinese: A functional perspective.* Taipei: Taiwan: Student Book Co., Ltd.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it. An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences. Acquisition, Processing and Use* (pp. 153-172). Amsterdam and Philadelphia: John Benjamins.

Verspoor, M., Schmid, MS., & Xu, X. (2012). A dynamic usage based perspective on L2 writing. *Journal of Second Language Writing, 21*, 239-263.

Wang, W. (2013). 对外汉语初级阶段语块构建研究 [An investigation on formulaic sequence development at beginning level for Chinese as a second/foreign language]. 北京大学 博士论文.

Wendel, J. 1997. *Planning and Second Language Narrative Production*. Unpublished Doctoral Dissertation, Temple University, Japan.

Wiggelsworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*(1), 56-106.

Wigglesworth, G., & Elder, C. (2010). An investigation of effectiveness and validity of planning time in speaking test tasks. *Language Assessment Quarterly, 7*, 1−24.

Williams, J. (2011). Is There Anything Special about Writing? Paper presented at the Annual conference of American Association of Applied Linguistics (AAAL). Chicago, IL.

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, HI: Second Language Teaching and Curriculum Center, University of Hawai'i at Manoa.

Wray, A. (2002). *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries.* Oxford: University Press.

Wray, A. & Perkins, M. (2000). The functions of formulaic language: an integrated model. *Language and Communication, 20*, 1–28.

Xiao, Y. (2004). L2 acquisition of Chinese topic-prominent constructions. *Journal of the Chinese Language Teachers Association, 39* (3), 65-84.

Yamashita, J., & Jiang, N. (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly, 44*(4), 647–668.

Yang, W. (2014). Mapping the Relationships among the Cognitive Complexity of Independent Writing Tasks, L2Writing Quality, and Complexity, Accuracy and Fluency of L2 Writing. Unpublished Doctoral Dissertation, Georgia State University.

Yao, T. & Liu, Y. (1997) *Integrated Chinese. Level 1 Part 1.* Boston: Cheng & Tsui Company.

Yuan, B. (1995). Acquisition of base-generated topics by English-speaking learners of Chinese. *Language Learning, 45,* 567-603.

Yorio, C. A. (1989). Idiomaticity as an indicator of second language proficiency. In K. Hyltenstan & L. K. Obler (Eds.), *Bilingualism across the lifespan* (pp. 55-72). Cambridge: Cambridge University Press.

Yuan, F., & Ellis, R. (2003). The effects of pre-task and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics, 1*, 1–27.

Yuan, F. (2009). Measuring learner Chinese in fluency, accuracy and complexity. *Journal of the Chinese Language Teachers Association, 44*, 109-130.

Yuan, F. (2010). Impacts of task conditions on learners' output in L2 Chinese narrative writing. *Journal of the Chinese Language Teachers Association, 45*, 67-88.

Zhan, X. (2012). 框架语块及其在对外汉语教学中的应用 [Formulaic frames and its implications in teaching Chinese as a second/foreign language]. 四川外语学院硕士论文.

Zhou, J. (2007). 语块在对外汉语教学中的价值与作用 [The status and implications of formulaic sequences in teaching Chinese as a second/foreign language].《暨南学报（哲学社会科学版）》第 1 期.

Zhou, Q. (2009). 对外汉语语块研究 [An investigation on formulaic sequences in Chinese as a second/foreign language]. 华东师范大学硕士论文.

Zimmerman, C. (1997). Historical trends in second language vocabulary instruction. In J. Coady & T. Huckin (Eds.). *Second language Vocabulary Acquisition: A rationale for Pedagogy* (pp. 5-19). Cambridge: Cambridge University Press.