Effects of Topic Familiarity and Linguistic Difficulty on the Reading Strategies and Mental Representations of Nonnative Readers of Chinese

Cecilia Chang
Williams College, Massachusetts

Abstract
The present article reports on a study that investigated the effects of topic familiarity and linguistic difficulty on the reading strategies and mental representations of nonnative readers of Chinese. Four passages – topic familiar/language easy, topic familiar/language difficult, topic unfamiliar/language easy, and topic unfamiliar/language difficult –were used. Forty American college third-year CFL (Chinese as a Foreign Language) learners were divided into four groups according to their familiarity with the reading topics, and were randomly assigned to read either the language easy or difficult passage. Subjects performed the think-aloud task to reveal their on-line processing strategies; the mental representations of what they read were reflected in their subsequent written recall of text content. Results suggest that subjects in all four groups predominantly engaged in local-level processing with the exception of two types of global-level processing: monitoring one’s comprehension and generation of inferences. While monitoring efforts were motivated by both topic familiarity and linguistic difficulty, inferencing events were primarily facilitated by topic familiarity. In addition, topic familiarity was also found to have a facilitative effect on the mental representations of the reading passages whereas no effects due to linguistic difficulty was found.

Two major characteristics associated with reading in a second language (L2) are unfamiliar topics and the foreign linguistic code. As learners advance in foreign language study, they are more likely to encounter texts with unfamiliar topics, such as cultural beliefs or social changes currently taking place in the target society. In contrast to vocabulary and expressions that reflect mostly day-to-day experiences in beginning- and intermediate-level textbooks, there are increasing amounts of formal and idiomatic expressions, as well as vocabulary that represent concepts that might be totally foreign to the readers of advanced-level textbooks. Understandably, challenges brought by unfamiliar topics and difficult language simultaneously can be
overwhelming to second language readers and severely affect their comprehension. The first step toward improving reading comprehension is to find out the types of strategies that readers use in response to the various types of difficulties they encounter during the course of the reading process. This paper reports on a study that investigated how forty adult nonnative readers of Chinese at the college third-year level processed four texts of different degrees of topic familiarity and linguistic difficulty—topic familiar/language easy, topic familiar/language difficult, topic unfamiliar/language easy, and topic unfamiliar/language difficult—and how the texts were represented in their memory as evidenced by recall.

Current models of reading explain the reading process as an active, constructive one in which readers actively interact with the text by using all available resources, both from information presented in the text and from their world knowledge. Meanwhile, comprehension activities take place at multiple levels—including a lower level where letters and words are recognized and sentences understood as well as a higher level where inferences are made to connect text propositions and to assimilate new information into readers’ preexisting knowledge (Kintsch & van Dijk, 1978). While automaticity in identifying letters and words and in syntactic processing contributes greatly to the smooth construction of a verbatim representation of the text (e.g., Dufour & Kroll, 1995; Bates & MacWhinney, 1989), it is largely dependent upon readers’ linguistic knowledge. Likewise, domain knowledge, or lack thereof, affects activation of a proper schema upon which inference generation largely rests, as readers make sense of the relationships between text propositions in creating a semantic representation of the text (Anderson & Pearson, 1988; Carrell, 1983, 1987, 1994; Hudson, 1982).

Automaticity in lower-level processing is important to smooth comprehension because speedy access to the lexical/semantic and syntactic elements in a text frees up cognitive capacity for higher-level processing such as generation of inferences and engaging in comprehension monitoring. Just and Carpenter (1992) suggest that during any cognitive activities only a limited amount of cognitive resources are available at any one time and that various cognitive activities compete for this limited amount of processing resources. Consequently, a reader must actively monitor the state of their comprehension and have a good sense about when and why comprehension breaks down in order to judiciously orchestrate all available cognitive resources. A reader also needs to determine how much attention should be paid to processing activities at different levels in order to comprehend a text efficiently. The unavailability of certain cognitive resources may push the reader to shift reliance to other resources that are available during the comprehension process. Accordingly, weakness in any one particular stage, higher or lower, could result in a heavier reliance on other knowledge sources (Stanovich, 1981, 1980; Hoover & Gough, 1990).

The ability to engage in comprehension monitoring thus plays a central role in successful reading comprehension and has been shown to the skill level of readers within both L1 and L2 readers (Hare, 1981; Garner, 1981; Steinberg, Bohning, & Chowning, 1991; August, Flavell, & Clift, 1984; Paris & Myers, 1981; Carrel, Pharis, & Liberto, 1989). Comprehension monitoring is, however, difficult to notice during an automatic-pilot state until a triggering event occurs that impedes the comprehension. Since much effort is required to ensure smooth progress of comprehension, conscious control of one’s cognition might be most noticeable in the “debugging state” (Brown, Armbruster, & Baker, 1986).
Compared to reading in a first language, L2 readers presumably face more difficulties caused by unfamiliar linguistic code and unfamiliar content, especially those that are culturally bound. In the following sections, general theories of the mental processes involved in reading comprehension will be presented before impacts of topic familiarity and linguistic difficulty on L2 reading are discussed.

**Mental Processes Involved in Reading Comprehension**

Researchers who have explored discourse comprehension generally contend that comprehension is a result of dynamic interaction between the reader and the text at multiple levels. At the most superficial level, the reader notices features of the linguistic code in which the text is written and then encodes them into memory. The memory traces of the resulting verbatim representation of the text, however, are rather unstable and fade rapidly. Beyond the surface code, the reader attempts to examine the relationships between text propositions in an effort to create a coherent textbase representation of the text, which captures the explicit meaning relations among elements within a sentence and across sentences in the text. In the process, a small number of inferences might be generated to establish local coherence.

While a more stable copy of the text information that can be held in memory longer, the textbase representation does not guarantee the comprehension that ultimately results in learning. As stated by van Dijk and Kintsch (1983, p.342), “Learning from text is not usually learning a text.” Such contention accounts for why the same discourse may result in two very different interpretations when read by two different people. It further explains how a student can memorize a computer programming text but still not be able to write a novel program. Observations like these suggest that readers use their prior knowledge, in conjunction with the propositional content of a discourse, to construct a coherent mental representation of the situation that goes beyond the explicit meaning of the text. The construction of coherence is the result of inferential processes that take place as the reader proceeds through the text. The resulting representations have an enduring effect in memory and have become known as situation models (van den Broek, 1994; Singer, 1994; Graesser, Singer, & Trabasso, 1994; Cote, Goldman, & Saul, 1998).

As such, in addition to the ability to identify words and parse sentences quickly to get at the information explicitly stated in the text, skillful readers generate inferences both to work out relations among text elements that are not clearly stated and to relate text information to their preexisting knowledge in order to construct new knowledge structures that contribute to learning. Though labeled differently by various studies investigating on-line activities during reading, three types of inferences have been identified as being likely to occur during text processing: explanations, associations, and predictions. The three kinds of inferences are quite functionally distinct. Explanations are backward-oriented and serve to unite the focal sentence with previous information, whether text-based or prior knowledge-based (Keenan et al., 1984; Myers et al., 1987; Trabasso & Magliano, 1996b). Associations occurred when readers elaborate text from the focal sentence; therefore they are concurrent (Trabasso & Magliano, 1996b; Myers & Duffy, 1990). Predictions are future-oriented, inferring future, causal consequences of a focal event (Duffy, 1986; Trabasso & Magliano, 1996b; Fletcher & Bloom, 1988).
Although all three types of inferences are likely to occur during the course of comprehension, not all of them are equally beneficial to text comprehension. Being crucial for establishing textual coherence, explanations are generated frequently during the course of reading and are found to contribute to the recall of both narratives and factual paragraphs (Trabasso & Magliano, 1996b; Chi et al, 1994). Empirical evidence has also indicated that readers are rather conservative in making predictive inferences, fearing wrong inferences about the consequences. When the readers have some prior knowledge of the content or there are enough semantic and linguistic constraints from the prior text, however, predictive inferences might be generated (van den Broek, 1994; Magliano, Baggett, Johnson, & Grasser, 1993). Similarly, associations are elaborative in nature and are usually not directly supported by previous text. Consequently, associations are likely to be irrelevant and erroneous. There is evidence that skilled readers are more selective with respect to generating associations than less skilled readers (Trabasso & Magliano, 1996b; Whitney et al., 1991).

**Impacts of Linguistic Difficulty and Topic Familiarity on Second Language Reading**

Since the late 1970s, substantial amount of energy devoted to second language reading research has led to the common belief that a certain level of second language linguistic ability must be achieved before readers can read effectively in the target language. Researchers who subscribe to this stance contend that inadequate command of the target language creates a threshold for effective transfer of L1 comprehension skills to L2 reading when confronted with a difficult or confusing task in the second language, and substantial comprehension skills in L1 cannot compensate for deficient L2 processing at the lexical and syntactic levels (Clarke, 1979; Cziko, 1980). Specifically, when language competence in an L2 is underdeveloped, word identification becomes sluggish. L2 reading research has demonstrated that less competent L2 readers rely more on the mediation of translation from L2 to L1 in order to access the conceptual representation of the L2 word than competent L2 readers (Dufour & Kroll, 1995). In addition, when reading a L2 which has an orthography different from that of one’s own, L2 readers spent most of their energy in lower-level processing (Everson & Ke, 1997; Horiba, 2000).

In addition to lexical access, speedy syntactic processing of non-proficient L2 readers is also dependent upon their L2 language competence. Studies that have investigated how syntactic factors influence comprehension show that different languages induced different approaches to sentence interpretation (Bates & MacWhinney, 1989, Chen, 1992). Readers of English rely more on word order for syntactic processing whereas readers of German are shown to rely largely on animacy and agreement contrasts. By measuring Chinese speakers’ reading times at various levels (including character-, word-, sentence-, and text-level), Chen (1992) demonstrated that subjects, when reading Chinese texts, paused longer at the end of sentences and at the end of each line in the physical text layout than they did at each character or word; these findings indicate that contextual information, instead of information embedded in words, was more relied on in comprehension of Chinese text. Because of the difference in approaches to sentence interpretation induced by one’s own language, non-fluent L2 readers have been shown to employ their L1 syntactic processing strategies in L2 reading given their limited experience in the
target language (Koda, 1993). Thus, the interference caused by L1 processing expectation in L2 sentence interpretation may tax non-proficient L2 readers heavily for cognitive capacity. As such, word- and syntactical-level processing requires enormous attention from readers and casts more demand on short-term memory (Potter, So, von Eckhardt, & Feldman, 1984). The overload on short-term memory in turn leaves little activation for high-level processing. Thus, non-proficient reading strategies of L2 readers are often characterized by a heavy focus on lower-level processing, which has been shown to short-circuit the transfer of L1 reading strategies to L2 reading (Clarke, 1988) and impede the construction of a coherent mental representation of the text they read.

Beyond the challenge brought by unfamiliar linguistic code, L2 learners are also faced with foreign content that is often associated with L2 reading. Researchers have argued that, whether we are aware of it or not, reading is in essence a process of assimilating new information presented in the text into old knowledge already stored in memory, or, our schemata. What is understood from a text is a result of the particular schema that is activated at the time of text processing (Anderson & Pearson, 1988; Carrell & Eisterhold, 1983). Studies adopted the schema-theoretic view in both L1 and L2 reading have reported facilitative effects of familiarity with both text content (content schemata) and with the structure or rhetorical patterns of the text (formal schemata) on improved inferencing and comprehension as measured by recall (e.g. Carrell, 1983; Lee, 1986). At the same time, however, unfamiliar content and structure impede the activation of appropriate schemata to accommodate incoming information, hence present difficult challenges to the readers in the reading process, resulting in various degrees of noncomprehension (Johnson, 1981; Hudson, 1982; Carrell, 1981; Ostler & Kaplan, 1982). Research has also shown that even when readers possess the appropriate schema they sometimes fail to activate it, and their comprehension suffers as a result (Carrell, 1983; Carrell & Wallace, 1983).

Compared to L1 reading, unfamiliarity with the content or structure of a text written in second language may be more detrimental to L2 readers as it brings extra challenges to the readers over and above the linguistic difficulties presented by a foreign language. Due to the fact that accessing of appropriate schemata in L2 reading generally depends initially on the reader’s ability to understand the foreign linguistic code, L2 competence is presumed to play some role in the activation of relevant schemata. Hudson (1982) provided preliminary insights into the interactive relationships between linguistic competence and use of schemata in L2 reading when he tested the effects of three instructional treatments—Pre-Reading (PRE), Vocabulary (VOC), Read-Test/Read-Test (RT)—on comprehension of ESL students of three proficiency levels (beginning, intermediate, and advanced). PRE was designed to explicitly induce schemata prior to reading; VOC provided students with definitions of vocabulary essential for comprehension of the texts; and RT allowed for readers’ self-reconciliation by using local context to reinterpret an instantiated, or partially instantiated, schema. Results from the study showed that PRE and VOC interventions were more effective for readers at the beginning and intermediate levels, whereas RT was more effective (though not statistically significant) than either PRE or VOC for readers at the advanced level. Together, these results suggest that the readers of lower-level proficiency were limited by their linguistic ability to instantiate coherent schemata on their own to help comprehension. In contrast, the advanced
readers—being equipped with better linguistic knowledge—were able to construct or modify the schemata without other interventions to aid their comprehension.

Barry & Lazarte (1995) also investigated the effects of prior knowledge and syntactic complexity on Spanish L2 readers’ comprehension as measured by recall. In their study, readers of high and low prior knowledge of Incan history read texts of three levels of syntactic complexity concerning the Incas—level 1, level 2, and level 3—and recalled what they read. Level 1 contained the essential ideas for the topic and served as the base reading passage whereas level 2 and level 3 contained different numbers of embedded clauses inserted in the base reading passage. Results from this study showed that while high-knowledge subjects recalled significantly more essential propositions in level 1 and level 2, their recall of essential information in level 3 was almost equal to those for low-knowledge subjects. Together, these results suggest that while prior knowledge facilitated recall of content for texts with lower syntactic complexity, the advantage of prior knowledge was cancelled out by high syntactic complexity.

Using the same data generated in their 1995 study, Barry & Lazarte (1998) investigated the effects of prior knowledge and syntactic complexity on the generation of three types of inferences in the recall protocols—within-text inferences, elaborative inferences, and inaccurate inferences. Within-text inferences were inferences made using information contained in the text, whereas elaborative inferences were generated with readers’ world knowledge. They found that high-knowledge readers generated significantly more inferences in general than low-knowledge readers. In addition, with the exception of inaccurate inferences, generation of inferences increased as syntactic complexity intensified. Specifically, for both high- and low-knowledge subjects, within-text inferences increased in level 2 and level 3 of syntactic complexity, but not in level 1, while elaborative inferences increased only in level 3 for low-knowledge subjects. Overall, however, elaborative inferences were generated less frequently than within-text inferences. Together, these results suggest that low-knowledge subjects constructed a representation that only reflected the surface code of the text with little evidence of inference generation. While there was a tendency to increase inference generation as text difficulty increased, the inference generation of low-knowledge subjects remained minimal, suggesting that they were not able to hold or retrieve information in their working memory from prior segments of the text.

Purpose of the Study

While Hudson (1982) and Barry & Lazarte (1995, 1998) provided valuable insights into the interactive effects of prior knowledge and linguistic competence on L2 readers’ comprehension as measured by different forms of comprehension tests, what remains unclear is how these two factors affect the shift between lower- and higher-level processing, presumably due to the insufficiency of cognitive resources. This is because these studies did not investigate the on-line cognitive processing attended to by the readers. Without a direct examination of readers’ on-line processing activities, we “understand something about reading comprehension, but we know little about how comprehension comes about” (Casanave, 1988). If we expect the results of our studies to have greater impact on actual classroom teaching, we need to provide insights into the effects as they are manifested during the comprehension process, not just the effects as an outcome. Such insights will strengthen the ability of language instructors to predict and diagnose comprehension breakdowns in actual
classroom teaching. These insights will be equally useful to textbook writers in compiling appropriate reading materials accordingly.

As only a handful of studies to date have dealt with the interaction between linguistic difficulty and topic familiarity in L2 reading, it is also important to examine to what extent the same findings from previous studies can be observed in studies with L2 readers reading a variety of different languages.

Accordingly, using the think-aloud technique to probe into the on-line cognitive activities of Chinese L2 readers as they read four types of passages—topic familiar/language easy (T-L-), topic familiar/language difficulty (T-L+), topic unfamiliar/language easy (T+L-), and topic unfamiliar/language difficult (T+L+)—this study is designed to investigate how reader’s familiarity with the reading topic and difficulty of the text language interactively affect the comprehension process in Chinese L2 reading, and if and how the effects differ from results of earlier studies with L2 readers of Spanish and English.

From the literature reviewed above, it should be safe to assume that L2 readers will employ largely lower-level strategies using local linguistic cues during the course of comprehension, to the extent that they are not fluent in the target language. Beyond this prediction, questions central to the interest of this study include which one of the two factors, language difficulty or topic familiarity, plays a bigger role in impeding or facilitating higher-level processing, such as generating inferences to construct a coherent mental representation of the text and to engage in metacognitive activities. An additional concern is in identifying the effects of topic familiarity and language difficulty on the resulting memory representations.

Accordingly, the research questions for this study are formulated as below:

1. What are the relative cognitive loads posted by the four types of reading passages (T-L-, T-L+, T+L-, and T+L+) and what are their effects on the allocation of cognitive resources, especially in terms of inference generation and comprehension monitoring, during the course of reading?
2. How do topic familiarity and linguistic difficulty affect the memory representations of the texts?

Method

Participants

This study targeted college CFL (Chinese as a foreign language) students at the third-year level (after 4 semesters of Chinese) as potential participants. There are two reasons why this level was of particular interest. First, after four semesters of Chinese study, students have typically learned enough characters (about 1,500) and knowledge of basic grammar to engage in processing of longer discourse without being bogged down too frequently by decoding difficulties. In spite of this, students at this level are still generally considered non-proficient learners, especially as far as
reading proficiency is concerned. Second, textbook language at this level starts to transition from everyday speech into more literary expressions, which is used frequently in formal correspondence and speeches. On the other hand, content of the lessons has become much more closely related to the culture and society of the target language and may be less familiar to students than that of typical first- and second-year level textbooks, which often draw on the personal experiences of learners (such as family/college life, traveling, or personal hobbies). As such, students at the third-year levels are faced with much more unfamiliarity of the text, in terms of both linguistic expressions and in terms of content, than their counterparts in first- and second-year courses. In other words, students at this level are not merely learning to read but also reading to learn.

As a measure to ensure that participants have relatively the same proficiency in the language at the time of this study, the study focused only on students who had little or no experience with Chinese prior to their formal study of the language. As a result, students who demonstrated better proficiency in speaking or reading Chinese either because they came from a Chinese-speaking family or because their native language employs Chinese characters as part of the orthographic system (such as Japanese and Korean) were excluded from this study.

In order to locate sufficient participants at the third-year level for this study under these constraints, study abroad programs, instead of regular Chinese programs at U.S. universities or colleges, were targeted as ideal sites for data collection for two reasons. First, due to high attrition rate, enrollments in upper-level Chinese language courses in regular U.S. colleges are usually unpredictable and are typically much smaller than those in lower-level courses. Also, students with prior knowledge in Chinese orthography (such as students of Chinese, Japanese, and Korean descent) are often found to constitute the majority of the student body in these courses, sometimes as high as 90%. This makes extremely difficult finding enough participants from one single school for quantitative studies beyond beginning and intermediate proficiency levels. This could be a contributing factor for the lack of empirical studies to date with learners of Chinese as participants at the more advanced proficiency levels. In contrast, accepting students from all U.S., Chinese study abroad programs, especially those that are competitive for admission, typically enjoy large enrollments (usually between twenty five and forth students) in third-year-level courses. Three study abroad programs were chosen for this study: Princeton-in-Beijing (P.I.B.), Associated Colleges in China (A.C.C.), and Inter-University Program for Chinese Language Study (I.U.P.). These three were chosen as study sites for this investigation because of their top-ranking status and the fact that they have comparable quality of instruction and criteria for admission, thus further ensuring the homogeneity of student backgrounds.

Data collection took place between the 3rd and the 4th week in summer at all three programs and again between the 3rd and the 4th week in the fall semester at A.C.C. and at I.U.P. Eventually, 55 students at the three study abroad programs who fit the selection criteria were initially recruited to participate in this study. After reviewing the data collected, participants who either performed extraordinarily well or

---

2 Chinese has been ranked by the Foreign Service Institute, which trains American diplomats, as one of the four most time-intensive languages to learn. An average English speaker takes 1,320 hours to become proficient in Chinese, compared with 480 hours in French, Spanish, or Italian. The participants in this study were estimated to have had at the most about 1100 learning hours at the time of the study.
badly on reading comprehension as measured by their recall protocols were removed from the subject pool, yielding a grand total of 40 subjects for this study.

Materials

In the two studies carried out by Barry and Lazarte (1995, 1998), student familiarity with the reading topic was determined by a pre-test treatment. Specifically, the high-knowledge group received instruction on the specific subject matter discussed in the test passage prior to the test while the low-knowledge group did not. This kind of pre-test treatment was determined to be infeasible for the current study since the study involved three study abroad programs; each has a tightly planned curriculum in place. Implementation of such pre-test treatment would constitute an interruption to the coherence of the curriculum and was therefore deemed undesirable. Consequently, this study relied on instructor judgment in determining the degree of topic familiarity. Note that accuracy and credibility of teacher judgment on student test performance has been widely investigated and documented in research studies that aimed at validating the tests themselves as well as arguing for greater teacher input in identification and placement procedures for different educational purposes (Morine-Dershimer, 1978-79; Coladarci, 1986; Luce & Hoge, 1978; Mayfield, 1979; Oliver & Arnold, 1978).

Subsequently, two prose texts—one with familiar topic and one unfamiliar—were first selected for this study from two Chinese textbooks to serve as the base materials by the author, who had over 15 years of experience in teaching Chinese to English-speaking college students in the United States at the time of this study. The two passages are also comparable to each other in length (378 and 377 characters respectively).

As explicated earlier, the content of third-year Chinese textbook lessons all have a certain degree of unfamiliarity to them, since this is a stage where students begin to read for the purpose of learning about the culture and the society of the target language. Consequently, the degree of familiarity of the reading text is defined here as the amount of background knowledge needed to comprehend the text.

The theme of the first passage deals with parental love, and the content of which is judged to be, relatively speaking, familiar to the students given its universal nature. In other words, students need little special knowledge about Chinese society in order to comprehend the text. The second passage deals with folk religions in Taiwan, and thus is assumed to be unfamiliar to most students (see appendix A).

For each passage, the language was manipulated first by the author to create two versions so that one would be more linguistically difficult than the other. The linguistic difficulty was created by replacing vocabulary items in spoken forms of Chinese with their counterparts in more literary forms, either in the form of lexicon or grammatical structure.

3 These two books were developed for third-year courses. The first, “Father’s Love,” was selected from the textbook “Talking of Chinese Culture” (1994) published by Beijing Language and Culture University Press; the second, “Taiwan’s Folk Religion,” was adapted from “Chinese Customs and Traditions” (1977) that was published by Cheng Chung Book Company in Taiwan.
Since there currently is no readability index for Chinese texts, this author then consulted the Modern Chinese Frequency Dictionary (1986) on all the points of manipulations in both texts, 23 between the T- passages and 25 between the T+ passages respectively (see appendix B), to make sure that the vocabulary items and grammatical structures used in the language difficult versions have lower frequencies than their counterparts in the language easy versions; a measure often used in assessing readability of English texts (Chall, 1995; Harrison, 1980; Johnson et al, 1994; Elley & Croft, 1989). Such manipulation of the texts resulted in four test passages: topic family/language easy (T-L-), topic family/language difficult (T-L+), topic unfamiliar/language easy (T+L-), topic unfamiliar/language difficult (T+L+), with the minus and plus sign represents low and high level of challenges respectively.

This author then consulted three professional Chinese instructors (who respectively had five, seven, and twelve years of experience in teaching Chinese to English-speaking college students at the time of the study) to further substantiate the degree of topic familiarity of the two base passages and suitability of the linguistic manipulations of each test passage. All three instructors concurred on the familiarity and unfamiliarity of the base passages. Furthermore, based on their feedback on the points of linguistic manipulations, modifications on the choice of literary forms were made twice more before the last versions of the test passages were finalized.

Procedure

Based on considerations to reader fatigue given the estimated time required to complete the two tasks involved – think-aloud and recall - in reading one passage (45 minutes to an hour), subjects were assigned to four reading conditions, T-L-, T-L+, T+L-, T+L+, and were asked to read only the one passage developed for that particular condition.

In order to achieve random assignment of subjects to reading conditions, several procedures were carried out. First, to tease apart those students who may have high prior knowledge in folk religions in Taiwan, such as graduate students with specialty in Asian religions, a questionnaire based on the Likert scale (1 being most unfamiliar and 5 most familiar) was first given to all students to evaluate their familiarity with different aspects of the Chinese society, including literature, history, political situation, and religion (see appendix C). The reason to include aspects of the Chinese society other than religion is to avoid clueing in the readers on the theme of the reading passage. Subsequently, only those students who indicated low familiarity in the religion aspect (1 or 2 on the questionnaire) were first selected to read the unfamiliar passages and were randomly assigned to the T+L- and T+L+ conditions, respectively. The rest of the subjects were then randomly assigned into the T-L- and T-L+ conditions.

To test if those subjects who were assigned to the T+ groups were weaker students in general in terms of their knowledge in the other areas of the Chinese culture/society, the author compared the means on the other familiarity scales (art, dialects, ancient history, modern history, and culture) for the students who indicated 1’s/2’s versus the 3’s/4’s/5’s on the religion scale using crosstabulations, with asymmetric Somers’D as the test of significance. Results showed no significant differences between the means of all the other familiarity scales for these two groups of students, suggesting that those students who were assigned to the T+ groups were not simply weaker students.
Efforts were also made to ensure relatively balanced representation of students from the three study abroad programs in each reading condition. The distribution of students from the three study abroad programs in the four reading conditions for final analyses is listed in Table 1.

**Tasks**

**A. Think-aloud**

To explore the on-line comprehension monitoring behaviors of the subjects, the think-aloud method was adopted for data collection. Think-aloud is a concurrent verbal report method that has been widely used in research on text comprehension both in L1 and L2 (e.g. Hosenfeld, 1977; Block, 1986, 1992; Zwaan & Brown, 1996). This method requires the students to verbalize any thought that comes to mind during the comprehension process (ranging from the types of difficulties they encounter to the strategies they use for comprehension) and to construct a meaningful representation of the text. It is widely accepted that think-aloud protocols directly tap into the type of cognitive activities readers do when they process text that would have otherwise been hidden for observation.

Before reading the test passages, students were given a different text to practice think-aloud techniques. A list of the things that people might report during reading was also shown to the students to further familiarize them with the think-aloud task (see appendix D). During the reading process, a red dot at the end of each sentence signaled students to pause and perform the think-aloud task. They were told that the goal of the reading task was to comprehend the meaning of the text, and that they should remember as much as they could about the content for they would later be asked to write down in paragraphs about everything they read, without referring back to the text. To avoid possibly incomplete reporting due to weak oral command in the target language, all think-aloud protocols were done in English, the native language of all participants.

**B. Recall**

After the think-aloud session, each student was asked to put away the passage and from memory write down in English the content of the passage they had read. They were instructed to recall as closely as possible the original text, both in terms of structure and in actual words used.

**Transcription**

Transcriptions of the think-aloud sessions were done by this author (a native speaker of Mandarin Chinese) and two research assistants who were native speakers of English with high proficiency in Mandarin Chinese. A sample of part of the transcribed think-aloud protocol can be found in appendix E. Statements in the protocols were then categorized into spreadsheet for data analysis. The categories for analysis of the think-aloud protocol are based on the study by Cote, Goldman, and Saul (1998) (with only minor revision). More specifically, think-aloud protocols were coded both for event frequency and event type.
Classification of Think-aloud Protocols

Number of events

As defined by Cote, Goldman, and Saul (1998), an event is a comment or set of comments on the same core sentences as well as the reading behavior associated with those comments. Two independent raters divided 5% of the protocols into events and agreed on 87% of the divisions. Disagreements were resolved in discussion, and the raters then coded another 5% of the protocols, with agreement this time reaching 93%. After resolutions of the disagreements, one of the raters then divided the remaining protocols into events.

B. Types of events

Each event was coded into one of seven major categories according to the nature of the comment. The event categories were as follows: paraphrasing, errors, monitoring activities, inferencing, sources of problems, repair attempts, and ignoring (see Table 2). Except for paraphrasing, inferencing, and ignoring, the other major categories all contain two to three subcategories, as will be explicated in the sections below.

Paraphrasing is where the student restated the statement using his/her own words. In the current study, paraphrasing is further defined as a correct translation of the focal sentence on the first try, thus will be indicative of the small amount of cognitive resources needed for processing.

The category of “error” encompassed both translation errors and decoding errors. An error event is defined as an event where a mistake is made without any follow-up verbalization from the student indicating that s/he knew or suspected that a mistake had been made and is viewed as a breakdown in monitoring effort. While correct translation of a phrase or sentence was treated as a type of paraphrasing event, translation errors, along with decoding errors, were combined and coded into an independent category.

Monitoring activities are events in which students verbalize either their current state of comprehension or the anticipated strategies required to ensure comprehension. There were several types of monitoring activities that occurred within the think-aloud protocols: confirming prediction or expectation (e.g., “That’s what I thought”), comments on text structure (e.g., “That’s the title”), noticing contradiction between text elements (e.g., “That doesn’t make sense”), and comments on one’s own behaviors (e.g., “I need to remember this” or “I’d better reread this part again”). These types of monitoring activities require the reader to step back and evaluate reading comprehension at a global level, paying attention to the meaning of the whole text and how ideas across the text are connected rather than restricting one’s attention to local coherence. Therefore, they are considered higher-level processing strategies. Due to the low rate of occurrence of the first two types of monitoring activities, the category of monitoring events in the final analyses consists solely of events where students noticed contradiction between text elements and where students made comments on their own behavior.

In this study, the inferencing category included only one type of inference: explanation; association and prediction were excluded from the final analyses due to
their extremely low frequency. Explanatory inferences are generated to approximate meanings of unknown characters, words, and phrases, as well as to establish connections between text propositions using one’s background knowledge or information from earlier parts of the text.

To reflect the types of difficulties students encounter during reading and the subsequent strategies they used to resolve these problems, two separate categories—sources of problems and repair attempts—were created. Six types of sources of comprehension problems were initially identified based on the think-aloud productions based upon the causal root of the misunderstanding: characters, words, phrases or sentences, relation of one sentence to another, relation of one sentence to either a paragraph or to the theme of the entire passage, and the theme of the passage. In the final analysis of the study, only three types of sources of problems (character problems, word problems, and phrase/sentence problems) were included, as there were too few cases of the other types to warrant further statistical analyses. Accordingly, only attempts to repair character, word, and phrasal problems were included in the final statistical analyses.

Lastly, cases where a student identified, decided to ignore, and hence made no attempt to solve a problem are coded as events of “ignoring.”

Two independent raters, the author and one research assistant, first coded the events in 5% of the protocols, agreeing on 83% of the event types. After resolving the disagreements in discussion, the raters then coded events in another 5% of the protocols, reaching a new agreement level of 91%. The inconsistencies were once again resolved in discussion. The author then coded the events in the remaining protocols.

Scoring of Recall Protocols

Scoring of the recall-protocols was based on a propositional analysis system put forth by Meyer (1985). Using a tree diagram similar to those used in linguistic analysis, the system portrays the hierarchical nature of text structure by reflecting the macropropositions, which serve to organize the relationships among ideas in the paragraphs, as well as the micropropositions, which deal with how ideas are organized into sentences and the ways sentences relate to each other within a text. The text structure can thus be used as a template for scoring recall protocols, where points are given for the presence of recalled idea units both in terms of lexical predicates as well as relationships.

The author and a research assistant scored the recall protocols. In preparation, the author and the assistant familiarized themselves with the derived text structure so as to gain quick access to the location of the information in the structure. They then read each protocol completely before actual scoring. Inter-rater reliability rate after scoring 5% of the protocols was 82%. Disagreements were resolved in discussion, and the raters then coded another 5% of the protocols, with the agreement reaching 92%. After resolving the disagreements, the author then scored the rest of the protocols.

---

4 This category includes only events in which a student identified but decided to ignore the problem without making any attempt to solve the problem. This is separate from events where students skipped the problem after unsuccessful attempts to find a solution.
6. Results

The dependent variable for all analyses was the proportion of any event type out of all total events. For major categories that do not have any subcategories (paraphrasing, inferencing, and ignoring), two-way ANOVAs (2 topic x 2 language) were used to identify differences among the four treatment groups due to topic familiarity and language difficulty. For the major categories that contain subcategories (error, monitoring, problem source, and repair) Two-way (2 topic x 2 language) repeated measures with Greenhouse-Geisser adjustment were used to assess differences due to topic familiarity, language difficulty, and subcategories of each major category. The between-subjects effects correspond to differences among groups on major category proportions; the within-subject effects correspond to differences in proportions of subcategories. If there were statistically significant differences in the major category, the Tukey HSD Procedure was used to identify specific differences among the four treatment groups. An alpha level of .05 was chosen as the significance level.

Table 3 shows the numbers of total events, mean proportions, and mean standard deviations of each major event category in the four reading conditions. While the total numbers of events provide a basic understanding of all cognitive activities induced in each condition, the different types of activities taken place will yield insights into the ways students allocated their cognitive resources and the challenges they faced during comprehension. The sections below will report the analyses of each event type.

For paraphrasing events, two-way ANOVAs showed significant effects of topic \([F (1, 36) = 13.84, p < .05]\), language \([F(1, 36) = 12.99, p < .05]\), and topic by language interaction \([F(1,36) = 4.83, p < .05]\). Taken together, these results showed that students who read the topic familiar passages paraphrased more than students who read the topic unfamiliar passages while students read the language easy passages paraphrased more than students who read the language difficult passages. In addition, the increase of paraphrasing events when reading the language easy passage was significantly more in the topic-familiar conditions than in the topic-unfamiliar conditions. Further analysis of mean proportions using Tukey HSD showed that students in the T-L- condition paraphrased significantly more [32%] than students in the T-L+ [14% ], T+L- [14% ], and T+L+ [9%] conditions. See Table 4.

Error

Repeated measures between-subjects effects of error events showed that the main effects of topic \([F(1, 36) = 23.89, p < .05]\), language \([F(1, 36) = 4.12, p < .05]\), and topic and language interaction \([F(1, 36) = 8.01, p < .05]\) to be significant. Students who read the topic unfamiliar passages made significantly more errors than students who read the topic familiar passages. Furthermore, in topic familiar conditions, students made more errors when the language was easy, whereas in topic unfamiliar conditions, students made significantly more errors when language was hard, so much so that students who read the T+L+ passages were found to make more mistakes by a big margin than students who read the other three passages [27% for T+L+ vs. 12% for T-L-, 10% for T-L+, and 17% for T+L-].
Within-subjects effects with translation errors and decoding errors as factors only showed an overall effect due to type of errors \([F(1, 36) = 8.32, p<.05]\), and no interaction of type of error with either topic or language was found. As such, students in all four conditions made significantly more decoding errors than translation errors. See Table 5.

**Monitoring**

For monitoring events, repeated measures between-subjects effects showed significant effects of topic \([F (1, 36) = 6.43, p<.05]\), language \([F(1, 36) = 21.92, p<.05]\), and topic by language interaction \([F(1,36) = 14.23, p<.05]\). Together, these results reflected that students who read the topic-familiar passages monitored more than their counterparts who read the topic-unfamiliar passages while students who read the language difficult passages monitored more than students who read the language easy passages. Also, the increasing effort for monitoring as language difficulty increases is much more substantial in the topic-familiar than in the topic-unfamiliar conditions. Additional analysis of mean proportions using Tukey HSD showed that students in the T-L+ condition monitored significantly more [15\%] than students in the T-L- [2\% ], T+L- [4\% ], and T+L+ [5\%] conditions.

Comparing types of monitoring activities, within-subjects effects with noticing contradiction and commenting one’s own behaviors as factors showed that there was a significant difference between the two overall \([F(1, 36) = 16.65, p<.05]\). In addition, interactions with language \([F(1, 36) = 24.83, p<.05]\) and topic by language \([F(1, 36) = 9.55, p<.05]\) were also present. As such, students monitored their own reading behavior more frequently [5\%] than they noticed contradictions between text elements [2\%]. Also, it was found that students monitored their own behavior more in language-difficult conditions than in language-easy conditions, but language difficulty made no difference on noticing contradiction. In addition, monitoring efforts occurred much more for commenting on own behavior in T-L+ [12\%] than any of other conditions [0\% for T-L-, 2\% for T+L-, and 4\% for T+L+] or for noticing contradictions (0-3\%) in any condition. See Table 6.

**Inferences**

For inferencing events, two-way ANOVAs showed significant effects of topic \([F (1, 36) = 9.84, p<.05]\), reflecting the fact that students who read the T- passages generated more explanatory inferences than students who read the T+ passages. Further analysis of mean proportions using Tukey HSD showed that students in the T-L- [10\%] condition generated significantly more inferences than students in the T+L- and T+L+ conditions [4\% and 4\% respectively], but not students in the T-L+ condition [7\%]. See Table 7.

**Problem Source**

For problem sources, between-subjects effects showed the main effect due to topic \([F(1, 36) = 4.31, p<.05]\) to be significant. Students who read the T+ passages noticed more linguistic problems than students who read the T- passages. Further analysis of mean proportions using Tukey HSD showed that students who read the T+L- passage noticed significantly more linguistic problems [42\%] than students who read the T-L- passage [31\%] but not students who read the T-L+ passage [37\%] and students who read the T+L+ passage [38\%]. On the other hand, within-subjects effects from repeated measures with character, word, and phrase problems as factors showed an overall effect due to types of problems \([F(1.66, 72) = 265.91, p<.05]\);
namely, character problems [27%] were articulated much more frequently than word and phrasal problems [both at 5% and 5%] across all conditions. No other interactions either due to topic, language, or topic by language were found. See Table 8.

**Repair**

Repeated measures between-subject effects of repair events showed a significant topic by language interaction [$F(1,36) = 4.98, p < .05$]. When reading the T- passages, students engaged in more repair when language is difficult [11% for T-L-, and 16% for T-L+], whereas when reading the T+ passages, such effort was more frequent when language is easy [18% for T+L-, and 13% for T+L+].

Repeated measures within-subjects effects with character, word, and phrase as factors were also carried out to see if students engaged in repair activities on one type of problem more than the others in any condition. In addition to a significant difference due to types of repair [$F(1.63, 72) = 34.37, p < .05$], a significant interaction with topic was also found [$F(1.63, 72) = 10.92, p < .05$]. Results showed that students engaged in phrasal repair [8%] significantly more than word repair [5% ], which in turn was significantly more than character repair [2%]. Also, there were significant differences between the ways students engaged in character vs. word repair and between word vs. phrasal repair due to topic familiarity. Specifically, students did more character repair in the topic familiar condition [2%] than in the topic unfamiliar condition [1%], whereas they did more word repair in the topic unfamiliar condition [8% ] than in the topic familiar condition [3%]. On the other hand, while efforts in word repair intensified when topic unfamiliarity increased [3% for T- conditions, and 8 % for T+ conditions], efforts in phrasal repair decreased when topic unfamiliarity increased [9% for T- conditions, and 7% for T+ conditions]. See Table 9.

**Ignoring**

For ignoring events, two-way ANOVA showed no significant effects due to topic, language, or topic by language interaction. The small proportions of ignoring events across all four conditions indicate that students in this study actively engaged in regulatory effort when a problem was detected, and rarely give up on a problem without attempting solving it first. See Table 10.

**Recall**

Results from two-way ANOVA on the effects of topic familiarity and linguistic on recall showed a main effect of topic, $F(1, 36) = 4.5, p < .05$. Students who read the topic-familiar passages recalled information significantly better than students who read the unfamiliar passages. In addition, between the L- passages and L+ passages, students recalled almost equal amounts of information, regardless of whether the topic was familiar or unfamiliar. See Table 11.

**Discussion**

Literature of L1 and L2 reading research suggests that insufficient prior knowledge of the reading topic and underdeveloped language competence both cast great demand on short-term memory and quickly deplete cognitive capacity. Effort in higher-level processing is thus severely obstructed; instead, lower-level processing is heavily relied on for comprehension. This study probes into how forty third-year level Chinese L2 readers allocated their cognitive resources when reading
four passages of varying degrees of topic familiarity and language difficulty (T-L-, T-L+, T+L-, T+L+), presumably as a consequence of the different cognitive demand posted by the two factors interactively. It is hypothesized that these readers, given their relatively undeveloped linguistic proficiency, would spend most of their energy in lower-level processing across conditions. Higher-level processing is most likely to take place when it is needed to compensate inefficiency in lower-level processing or when there are enough cognitive resources left after local linguistic elements have been reconciled. As such, the shift between lower-level and higher-level processing should be indicative of the relative degree of demand on cognitive capacity posted by topic familiarity and language difficulty in each condition, with larger proportion of lower-level processing strategies indicating heavier cognitive load. Other indirect indications to the respective cognitive demand in each condition may manifest themselves in how much the readers are able to understand a sentence upon the first try (paraphrasing); how much they fail to notice a mistake they made (errors); how frequently they give up a detected problem (ignoring), presumably as a result of overload on their cognitive capacity; and how well they recall the content of the passage. In addition, based on the findings from Barry and Lazarte (1995, 1998), it is hypothesized that students who read the familiar passages will have better recall than students who read the unfamiliar passages. Also, for both the familiar and unfamiliar passages, students who read the linguistically easy passage will have better recall than students who read the difficulty passage.

Think-aloud (concurrent introspection) and written recall were adopted as the methods of data collection. It is commonly accepted that comprehension processes can be inferred from what is reported in think-aloud protocols as they reveal the content of information available, or in this study, unavailable, to the reader’s immediate awareness of short-term memory and the processes that require the reader’s conscious attention (Ericsson & Simon, 1980, 1984). Recall protocols are also widely acknowledged for their utility in reflecting reader comprehension without clueing in the reader on the type of information expected.

Effects of Topic Familiarity and Language Difficulty on the General Processing Patterns and on the Relative Cognitive Loads of Each Condition

In terms of allocation of cognitive resources between lower-level and higher-level processing, our hypothesis that lower-level processing would dominate the various cognitive activities that take place among the students was confirmed. Together, local-level monitoring and repair effort (i.e., noticing and reconciling character, word, and phrasal problems) constituted about half of the cognitive activities that took place in all four conditions (42%, 53%, 60%, and 51%, respectively). In contrast to the high occurrence of lower-level processing activities, higher-level processing activities (i.e., global-level monitoring and inferencing events) not only took place rather infrequently but also with large differences among readers as well as among types. For example, while four types of global monitoring activities were found throughout the think-aloud protocols (i.e., confirming prediction or expectation, commenting on text structure, noticing contradiction between text elements, and commenting on one’s own behaviors), the former two types of monitoring activities were excluded from the final statistical analyses due to their low occurrences. While the latter two were included in the final statistical analyses, their
small proportions compared with other event types suggest that they were not in the forefront of the mind of the readers.

With regard to inference generation, students in this study only generated limited amount of explanatory inferences; predictive and associative inferences, on the other hand, were either not generated in this study or were generated with extremely low frequency. It appears that these students did not actively resort to their world knowledge to solve local linguistic problems nor did they try hard to go beyond the boundaries of the focal sentence for information that could help them construct a bigger context for the text they were reading.

These findings indicate that the majority of the students in this study centered their energy on attending to the linguistic problems appeared in the focal sentence and devoted little capacity for anything else that may take their attention away from the sentence they were reading. They seemed to have spent little energy on other components crucial to successful text comprehension, such as noticing text structure and establishing intersentential coherence. They also avoided making irrelevant associations or possibly erroneous predictions, either due to lack of cognitive resources or as a result of excising extra caution. While they rarely gave up on any comprehension problems encountered, their approach to solving the problems has been to rely on cues in the immediate context, instead of utilizing information already read or from their world knowledge.

These findings once again confirm the contention that L2 readers’ text processing operations are primarily influenced by demands of language processing in that lower-level processes are prioritized at the expense of higher-level processes when the demand for processing resources is greater than the supply in non-fluent L2 comprehension (Potter, So, von Eckhardt, & Feldman, 1984; Stanovich & West, 1981).

With specific regard to Chinese L2 reading, this study has shown some interesting processing patterns that could have resulted from the orthographic and linguistic features of Mandarin Chinese. In particular, students in each reading condition articulated problems with unclear characters much more frequently than with unclear words or phrases, but they engaged in phrasal-level repair significantly more than they engaged in character- or word-level repair. Given that character boundaries, not word boundaries, are marked by space in Chinese orthography, Chinese characters have become basic perceptual units in Chinese reading and function as coding units similar to those of alphabetic words (Chen, 1984, 1986, 1987). The fact that students in this study articulated more character problems may be treated as a direct consequence of this particular feature in Chinese orthography. However, the reliance on phrasal context for comprehension requires careful interpretation since such reliance was both found in Chen (1992) by native Chinese readers and in the present study by non-native Chinese readers. As such, while it is possible that the students in this study used contextual clues to compensate for their deficiency in lexical access in an effort to conserve cognitive resources, such a

---

5 In contrast to Indo-European languages, the Chinese language possesses a number of unique orthographic and linguistic properties, summarized by Chen (1992) as follows: Chinese characters represent lexical morphemes, and they vary in construction complexity; Chinese words can be formed by one or more characters; character boundaries, but not word boundaries, are indicated by spaces; and Chinese words generally have no inherently marked lexical categories nor inflectional markings.
strategy could also have been induced by linguistic features associated with Chinese
morphology.

In terms of the relative degree of cognitive demand posted by the four
passages, we should be able to safely assume that the T-L- and the T+L+ passages
posted the lightest and the heaviest cognitive loads on the students respectively. Being
best able to translate/understand what they read upon first try (paraphrase), students in
T-L- eliminated the need in engaging in activities in subsequent events; such activities
included noticing problem sources or making necessary repairs, making the total
number of cognitive events of this reading condition the smallest of all four
conditions and resulted in the strongest recall performance among the four groups. In
contrast, the comprehension challenges brought by both unfamiliar topic and difficult
language in the T+L+ condition resulted in significantly reduced proportions of
paraphrasing and inferencing. Without proper prior knowledge of the reading topic
and linguistic knowledge of the L2, students in this condition had a harder time
understanding a focal sentence upon first try and their ability in using information
from sources beyond the focal sentence for inference generation seemed to be heavily
impeded. Additionally, being inundated with cognitive demand coming from
processes at multiple levels, these students appeared to be far less efficient in
detecting problems, and the decoding and translation errors they made along the way
increased significantly as a result. Jointly, these processing constraints might have also
contributed to the recall by students in this condition being the weakest among the
four groups.

Compared with the T+L- passage, the T-L+ passage seems to have posed
similar but perhaps lighter cognitive demand on the students. Although initial
observation of the numbers of total events in the two conditions indicated that
students in the T-L+ condition engaged in more cognitive activities than students in
the T+L- condition, further examination showed the difference was caused by an
increase in global monitoring, a higher-level processing activity, in the T-L+
condition. At the same time, there is indication of a higher occurrence of events
associated with local linguistic difficulty in the T+L- conditions. In addition to the
slightly higher proportions of errors and repair attempts, students in T+L- condition
were also found to have articulated problems with local linguistic elements
significantly more than students in the other three conditions. These observations,
together with the fact that students in T+L- condition produced much weaker
performance on recall than students in the T-L+ condition, lead us to the conclusion
about the respective cognitive loads in the two conditions.

**Effects of Topic Familiarity and Language Difficulty on Higher-Level Cognitive Activities**

In this study, students’ insufficient linguistic competence in L2 was found to
have restricted student comprehension strategies mostly to the lower-level in general.
These readers did engage in higher-level processing, however, presumably when they
had more cognitive resources at their disposal or to compensate for deficiency in
lower-level processing. The two higher-level cognitive activities in this study are
monitoring and generation of global inferences. So, what are the roles of topic
familiarity and language difficulty in impeding/facilitating these higher-level processes?

In terms of monitoring activities, linguistically difficult passages motivated students to monitor their comprehension more than linguistically easy passages, corroborating the findings by Brown and Baker (1984) in that monitoring effort is most noticeable when there are greater challenges present. Comparing the increase of monitoring activities between the L- and L+ passages in both the T- and the T+ conditions, we observed that students who read the T-L+ passage not only monitored significantly more than students who read the T-L- passage, but also more than students who read the T+L- and who read the T+L+ passages. In addition, while there was an increasing trend in monitoring behaviors between the L- and the L+ passages in the T+ conditions, the increase was far less substantial that that in the T- conditions. It is hard to know whether students who read the T+L+ passage simply did not engage in substantially more monitoring activities than students who read the T+L- passage, or if they indeed monitored more but were too flooded with other tasks to verbalize their effort. Whatever the case may be, these results together suggest that student motivation in engaging in monitoring activities was generally strengthened as they were faced with a linguistically more difficult text. It is, however, when the topic of the passage was also familiar that the students appeared to have extra capacity to verbalize their effort more fully.

With that being the case, however, we need to take into consideration two other factors, namely, the nature and the relative occurrence of the two monitoring activities involved, in order to get a better picture about how topic familiarity and language difficulty affect these students’ monitoring effort. Comparison of the occurrence of the two types of monitoring activities showed that students commented on their own behaviors significantly more than they noticed contradictions between text propositions or between text propositions and their own expectations. While students commented on their own behaviors more when reading the L+ passages than in the L- passages, language difficulty did not affect the way students noticed contradictions—energy spent on noticing contradiction remained equally low (2%) in L- and L+ conditions.

Comments on one’s own behavior (e.g., “I need to reread this part” or “Oh, I got it”) are usually generated as a response to a concurrent situation, hence, readers are not required to search for information stored in their short-term memory as they would in noticing contradictions. Consequently, commenting on one’s own behaviors is assumed to consume much less cognitive resources than noticing contradictions. The small overall proportion of energy devoted to noticing contradictions (2%), together with the predominant lower-level strategies among all cognitive activities that took place, suggests that the cognitive resources of the students in this study were heavily taxed for linguistic processing and there was little activation left for either real engagement of or articulation of higher-level comprehension monitoring; when cognitive resources did become available (seemingly due to being familiar with the reading topic), there seemed to be only enough capacity for undemanding monitoring.

As stated earlier, students in this study allocated limited amount of cognitive resources to inference generation, indicating that they did not actively use information beyond the focal sentence for comprehension. Further analysis showed that students who read the T- passages generated more inferences than students who read the T+ passages (8% in T- conditions and 4% in T+ conditions). As such, student ability to
use background knowledge and information beyond the focal sentence for comprehension was influenced by whether or not they had the necessary pre-existing knowledge about the reading topic. When a proper schema could not be activated, little information could be retrieved from long-term memory to accommodate incoming information. Not surprisingly, inference generation decreased when students were faced with a topic-unfamiliar text.

**Effects of Topic Familiarity and Language Difficulty on Recall**

Regarding student comprehension as measured by recall, the study showed that, first, students who read the topic-familiar texts were significantly better at recalling information than students who read the unfamiliar texts. This part of the results confirmed our prediction about the recall performances by the four groups of students and corroborated well with previous studies (Hudson, 1982; Carrel, 1983, 1984, 1987) on the facilitative effect of prior knowledge on comprehension. Second, regardless of topic familiarity, students reading the L- passages and students who read the L+ passages recalled almost equal amounts of information from the texts. This was a little surprising as, based on the results from Barry and Lazarte (1995), we were expecting to see a facilitative effect of high prior knowledge on recall of information for the language-easy texts.

The discrepancy may be due to the different nature of linguistic difficulty in the Barry and Lazarte study and the current study. Specifically, in Barry and Lazarte (1995), different numbers of relative clauses were embedded into the base text, creating three levels of syntactic complexity. With each increased level of syntactic complexity, not only was the length of the text increased, but the flow of information was also slightly altered, requiring more effort in keeping incoming information in line with the theme of the reading text. In other words, the embedded relative clauses ultimately increased the number of propositions in the reading texts and increased the overall length of the texts. In the current study, however, linguistic difficulty was created by replacing colloquial expressions with their counterparts in formal literary forms. Therefore, the language difficulty was mainly manifested on the lexical level and no extra proposition was added. Consequently, findings from the two studies showed a different effect of interaction between high prior knowledge and linguistic difficulty on recall of text information. It seems, thus, that linguistic difficulty created on the lexical level did not result in a significant difference in cognitive overload to have any effect on student recall of text content, regardless of familiarity of topic.

**Limitations, Conclusions, and Implications for Classroom Instructions and Future Research**

The current study is a first attempt at investigating the reading behaviors of high intermediate CFL readers at the discourse level, and suffers from some of the same limitations encountered in other studies that employ statistical analysis. First, the relatively small number of participants in each group limited the generalizability of the findings of the current study. Second, the between-subject design, while eliminating concerns about reader fatigue, opened the door to questions concerning effects attributable to individual differences among participants, such as L1 reading ability and motivation. At the same time, despite being widely accepted for its ability to probe into the comprehension processes more directly, the think-aloud method has been noted for either overestimating or underestimating the amount of information that readers maintain across sentences, often as a result of the pragmatics of the
reading task (Long & Bourg, 1996). Consequently, the conclusions of this study must be seen as limited and suggestive.

Despite these limitations, however, the current study yielded results that corroborate findings from earlier studies (Hudson, 1982; Barry and Lazarte, 1995, 1998) in many ways. To begin, the processing operations of nonproficient Chinese L2 readers, just like their counterparts in Spanish and in English, are still largely constrained by their L2 language competence, as seen in the large proportion of lower-level processing activities that took place regardless of the type of texts they read. Their inefficient access to the semantic and syntactic characteristic of the text cast enormous overload on the cognitive capacity, which further hindered their effort in higher-level processes in general, suggesting that they engaged in text-based processing more frequently than at the situational level. However, similar to previous studies, a facilitative effect of readers’ preexisting knowledge of the reading topic on comprehension was also observed in this study.

Since the materials used in this study were developed primarily through the judgments of language instructors, these corroborating results show that instructors’ judgments on test materials are equally as useful in predicting student reading performances as other experimental manipulations used in previous studies. Language instructors can feel more confident about making predictions of their students’ reading performance, and design strategic curriculum to improve their proficiency.

This study’s investigation of students’ on-line shift between higher- and lower-level processing also enriches the current knowledge about the effects of topic familiarity and language difficulty on L2 readers’ processing activities. Specifically, we gained insights into the relative cognitive loads posted by texts of varying degrees of topical and linguistic difficulties. These results are informative to language educators in developing and sequencing reading materials for their students. When developing thematically-based curriculum that aims at “reading to learn,” L2 instructors and material writers should bear in mind the cognitive load posted jointly by unfamiliarity of the reading topic and language difficulty. It may be desirable for the texts to be arranged according to their topic’s familiarity to students, introducing the more familiar ones first, followed by those with less familiar topics. Once the texts are sequenced according to their familiarity to students, they can be further arranged so that the language difficulty of the texts increases along with the unfamiliarity of the reading topic. Such a sequence allows students to progress from texts that are cognitively less demanding to those that are more demanding.

The facilitative effects of preexisting schema (owing to familiarity with text topic) on comprehension found in this study reconfirms the importance of strategic incorporation into L2 reading curriculum means of providing and activation of appropriate schema. The activated schema will act as an advanced organizer to facilitate the integration of new and old information. The forms of the advanced organizers and the amount of classroom instruction to be devoted to such practices, however, may depend on several factors. First, advanced organizers have been found to be more beneficial to subjects with lower proficiency than to subjects with more advanced proficiency, who presumably were not particularly in need of such facilitation, given their better-developed linguistic knowledge in L2 (Hudson, 1982; Adams, 1982; Mueller, 1980). Further, evidence from research in L2 acquisition has shown that learners learning a second language that is typologically similar to their
native language progress more rapidly than learners learning a second language that is typologically different from their own (Wood, 2004). Accordingly, it is quite plausible that the degree to which postintermediate readers would benefit from the advanced organizers may be jointly decided by a number of factors, including their proficiency levels and the distance between the target language and their native language, both morphologically and syntactically. Future studies are needed in this regard to help classroom instructors make informed decisions on the appropriate amount of time to be allocated between activating readers’ preexisting knowledge about the text and other classroom tasks based on the needs of their students at different proficiency levels.

Finally, in light of the preponderant patterns of lower-level processing among the CFL L2 readers in this study, there is no doubt that continuous attention should be paid to developing strong linguistic ability so that they can improve on the flexibility of their strategy use and more effectively utilize cognitive resources at multiple levels. At the same time, the lack of global-level processing activities among these readers signifies a call for classroom instruction in which strategies to raise awareness of various cognitive and metacognitive approaches to comprehending a written text are promoted.

BIBLIOGRAPHY


APPENDIX A READING PASSAGES USED IN THE CURRENT STUDY

Passages used in the topic familiar conditions (T-L- and T-L+) -- "Father's Love"

A. 父親的愛 (T-L-)

我父親年輕的時候是個軍人。母親常說他從來沒抱過孩子，從來沒參加過家長會。在這麼多的“從來沒”中，父親常不知道說什麼好。也許父親真的不太關心我們。可是有一次，父親突然說，我出生那天，他高興的不得了，跑到外面買了個鮮紅的鬧鐘。後來，再聽到那些“從來沒”，我眼前馬上就會看到那枝鬧鐘。

我工作後，有時晚上很晚回家，會發現父親站在黑暗的街口等我。日子久了，一直到今天，雖然我早就不習慣父親的保護，沒有自己的家，甚至沒有自己的黑路上，還會有一種被人關心的溫暖感。

父親很喜歡文學，但他從不寫作。在我最困難的時候，父親淡淡的說，你可以試著把你的感受寫下來，所以我就開始寫。後來那些想法一個一個成了文章發表了，而且都被父親收集起來了。有的時候看到雜誌廣告，父親會算好出版的日期去買。他一遍一遍的讀，熟悉我所寫的每個字。我深深的知道父親為我感到驕傲，所以我更努力。有一次，我告訴父親我已寫了一百多萬字了，他沉默了一會兒，說“別拼命寫”。這就是我的父親。

(378 characters)

B. 父之愛 (T-L+)

我父親年輕時是個軍人。母親常說他從來沒抱過孩子，從來沒參加過家長會。在眾多的“從來沒”中，父親常不知如何是好。或許，父親真的不太關心我們，然而有一次，父親突然說，我出生那天，他高興的不得了，跑到外面買了個鮮紅的鬧鐘。後來，再聽到那些“從來沒”，我眼前頓時就會出現那枝鬧鐘。

我工作後，有時晚上很晚回家，會發現父親站在黑暗的街口等我。多年來，直到如今，有時夢見，走到無人的黑暗中，仍會有一種被人關心的溫暖感，那種我早已習慣了父親的保護，有了自己的小家。

父親很喜歡文學，但他從不寫作。在我最困難的時候，父親淡淡的說，你可以試著把你的感受寫下来，所以我開始寫。後來那些想法一個一個成為了文章發表了，而且都被父親收藏起來了。有的時候看到雜誌廣告，父親會算好出書的日期去買。他一遍一遍的讀，熟悉我所寫的每個字。我深深的知道父親為我感到驕傲，所以我更努力。有一次，我告訴父親我已寫了一百多萬字了，他沉默了一會兒，說“別拼命寫”。這就是我的父親。

(362 characters)

Translation of the passages "Father's Love"
My father was a soldier when he was young. My mother often said that he never held the babies, never attended the parents meetings, ... Father would not know what to say to these complaints. Perhaps, he really didn't care about us. However, one time, father suddenly said to me that he was so excited the day I was born that he went out and bought a red clock. Later on, whenever I heard mom's complaints, the image of the red clock would appear in front of my eyes immediately.

After I started working, I would find father waiting for me on dark streets when I cam home late. Up until today, despite the fact that I have formed my own family and left father's protection, I would have a very warm feeling of being cared for when I walk on a dark street.

Father loves literature, but he never writes. During my most difficult time, he said to me: "You should try to write down your thoughts." Therefore I started writing. Later on, these thoughts turned into articles and were published one by one, and they were all collected by my father. Sometime, father would count the days to buy the magazines that publish my articles. He would read the articles again and again until he would almost memorize every word I wrote. I know deep in my heart that he is very proud of me, therefore I work even harder. One time I told him that I had written over one million words, he was silent for a while, and then he told me: "Don't over work." That's my father.