Breaking the Camel's Back: Cognitive Load and Reading Chinese

Jeffrey J. Hayden
University of Hawai'i at Mānoa

1. INTRODUCTION.

In the very first issue of the Journal of the Chinese Language Teachers Association (JCLTA) in 1966, Professor John DeFrancis published the article "Why Johnny can't read Chinese". Perhaps its prominent placement as the vanguard article in that first issue of the Association almost 40 years ago was a foretelling of how it was to become a seminal and often-cited paper in the field of Chinese language teaching, with particular regard to the teaching of reading.

Not until relatively recently has reading comprehension been looked at in terms of cognitive processing of extended text, and still very little research has been done with reading second/foreign languages in this regard. The main purpose of this paper is to try to examine the mentałcognitive processes involved during the reading of novel text as displayed on a computer screen. This paper also attempts to look at the role working memory appears to play among readers of Chinese.

2. Reading Chinese as a Second/Foreign Language (CL2/CFL)

2.1. The Burden of Orthography

One of the main reasons often cited for causing so much difficulty among CL2/CFL students learning to read Chinese are the characters. In fact, five of Moser's (1991) nine reasons why Chinese is so hard deal with are:

1. Because the writing system is ridiculous.
2. Because the language doesn't have the common sense to use an alphabet.
3. Because the writing system just ain't very phonetic.
4. Because you can't cheat by using cognates.
5. Because even looking up a word in the dictionary is complicated.

(Moser, 1991, pp. 60-65)

There are generally three main concerns when discussing the instruction of Chinese characters: retention, recognition, and production. Retention deals with the "tricks", often mnemonic, that students use to keep the characters in memory. Recognition implies that students have managed to successfully embed a character or word into more long-term storage to the extent that at the least they will be able to provide a pronunciation or a meaning associated with it. Finally, production means that not only have students retained the character or word and can recognize it, but that they can recall and more-or-less faithfully reproduce a character in the right situation.

2.2. Character Identification

A large number of studies looking at the processing of the Chinese written language have focused on individual characters (Chen, 1986; Feldman & Siok, 1999; Hayes, 1987; Ke, 1996, 1998; Perfetti & Tan, 1998; Sergent & Everson, 1992; Sun, 1993; Tzeng, Hung, & Garro, 1978; Wang & Thomas, 1992; Zhou & Marslen-Wilson, 1999). Key among these studies has been the issue of to what degree the effects of complexity, density, and frequency play in processing the information presented by a character. Frequency can either be 'frequency of occurrence' (how often a character is used in the written language in general based on statistics gathered from corpus studies) or 'frequency of exposure' (how often a character is repeated in a language textbook) (Sergent & Everson, 1992). In either case, high-frequency characters are used more often and are thus more common than low-frequency, or rare, characters. Complexity differs from density in that complexity by definition is more generally limited to number of strokes while density refers to the proportion of space filled to construct a character. As Yang and McConkie note, "additional complexity can arise in a number of different ways: by having non-decomposable characters with more strokes, or by having characters with additional components arranged in different ways" (1999, p. 218). For example, Figure 1 shows three pairs of characters where each pair has the same degree of complexity, yet differ in density:

6 strokes	8 strokes	10 strokes
回 好	京 事	笑 能

Figure 1. Visual difference between complexity and density in Chinese characters.

There have been a few studies that have looked specifically at the three factors of frequency, complexity, and density in character identification.[1] Hayes (1987) looked at the nature of character complexity as it affected character identification. He used a character recall task that asked subjects (17 advanced non-natives and 17 natives) to match characters they believed were shown during a four-second display on an answer sheet provided. He divided 105 characters into four groups of varying complexity: Low (1-5 strokes), Medium (6-10 strokes), High (11-15 strokes), and Complex (15 or more strokes) (p. 49).[2] His results showed that accuracy of recall was not affected by character complexity.

Sergent and Everson (1992) used a character naming task to look at textual frequency and complexity in terms of both speed and accuracy in two levels of non-native learners of Chinese, beginning (n=17) and advanced (n=5). The three independent variables were instructional level (beginning, advanced), frequency (low frequency [<35.1] and high frequency [>35.1], based on occurrences in the language textbook), and density (low density [<7] and high density [>7], based on number of strokes). The two dependent variables were vocal latency and accuracy. They observed that advanced readers were faster and more accurate in the character naming task than the beginners. Sergent and Everson do note, however, that

> "The longer reaction times for denser characters are congruous with the findings of Hayes (1987b) that non-natives were more likely than natives to attend to the visual aspects of characters when holding the character in short-term memory. The longer reaction times for the more visually complex characters indicate that non-natives do not process visually complex characters as rapidly as visually non-complex ones because they attend to the additional visual features of more visually complex characters. (1992, p. 41)

---

[1] Chen (1986) also notes that display size seems to be "an important factor in character/word recognition" (p. 9).

[2] I think there is a potential problem with these groupings, however; one that if corrected for might undermine his conclusion. If we look again at Figure 1, notice that two of the three pairs of characters represent the low and high ends of his medium group. This would mean that both 回 and 能—characters of obvious and strikingly dissimilar densities—would be considered as being of equal complexity. I believe there is even a density distinction between 事 and 笑 , where 事 could be identified as being more complicated than 笑 in a timed experiment even though 事 has fewer strokes.

Recently, Xiao (2002) looked at the effect of complexity again. For her study, she defined her complexity levels as low (1-6 strokes), mid (7-11 strokes) and high (more than 12 strokes). Her data came from post-lesson vocabulary quizzes of 34 first semester non-native students. The vocabulary quizzes were comprised of three sections: character recognition, character production, and dictation. Like Sergent and Everson (1992), Xiao found that character complexity affected recognition and production but hypothesized that complexity effects diminish as students progress to higher language levels and incorporate chunking in the form of radicals and phonetic components as opposed to focusing on strokes at the beginning of Chinese character learning.

Finally, though it only deals with recognition, a recent study by Wang, Perfetti, and Liu (2003) corroborates what has been discovered in the previous studies above in that increased frequency of exposure is strongly correlated with faster response times and more accurate identification of characters in a naming task. Not surprisingly, since Chinese characters do not indicate tone, the 15 beginning-level CFL participants (composed of two dissimilar groups) performed poorly on a tone matching task. They conclude that their findings indicate a primacy of perceptual learning among students whose native written language is alphabetic (e.g., English) in learning a non-alphabetic written language such as Chinese.

2.3. Word Recognition

In order to identify whole words, it is argued that beginning readers must first decode their component parts. Experienced readers, it is argued, do this mainly with infrequent, unfamiliar characters and words. For less frequent and not-quite-yet-automatized lexical items, this typically involves the assignment of sound first, then accessing meaning. For English and other alphabetic languages, word identification requires the ability to construct meaning from varying strings of letters, which are representations of sounds. For Chinese, a morphosyllabic language (DeFrancis, 1989), prelexical identification of words seems to rely on the ability to compile meaning from jumbles of seemingly random strokes constrained within a limited square space.

Another issue is the question of unknown vocabulary density (Hsueh-chao & Nation, 2000). The greater the extent of one's knowledge of vocabulary the more likely one's ability to read and comprehend a novel text. Of course, DeFrancis (1966) notes that even with a repertoire of 1200 individual characters, which represented 91.3% of characters in Chen's 1928 data (as cited in DeFrancis, 1966, p. 10), "knowing a character does not necessarily give a student access to the combinations of which it is a part" (p. 11). As Hayes (1990) argues, without a solid foundation of vocabulary (characters alone and in combination) as represented by a "memory network of Chinese" (p. 38), the ability to even exercise a top-down, conceptually-based, knowledge-driven approach to an extended piece of novel text is crippled with only limited potential for success at best. Everson (1994) reminds us that when readers are forced to grapple with a foreign language text, as opposed to reading it more or less fluently, they must "divide up their cognitive resources in a strategy termed 'code switching'" (p. 6).

2.4. Working Memory

Working memory (WM) is by all accounts and above all else understood as a workspace. This workspace is recognized as temporary in nature, as well as being cognitively based. The WM system is said to be limited in capacity, varying from individual to individual, and even subject to moods and emotions. Working memory, unlike long-term memory (LTM), is seen to hold onto the materials it is working with only for short periods of time (see Miyake and Shah, 1999, for an extended definition of working memory).

Working memory processing includes the encoding (e.g., phonological and semantic codes for language) of information for either immediate processing or storage for future processing, maintenance of information primarily through rehearsal mechanisms, and retrieval of information either currently in the working memory buffer or information that has been placed into LTM storage.

Looking at the data provided by Ke (1996), there seems to be some evidence suggesting that Chinese characters place high working memory processing demands on beginning learners of Chinese characters. Further, over time and with relatively constant exposure and use, lower-level processing does seem to give way potentially allowing for higher-level processing to take place more efficiently. Thus, both familiarity and frequency of character components, characters, and words directly influence the load placed on short-term memory (STM), processing demands placed on working memory (WM), and storage and recall efficiency of long-term memory (LTM).

2.5. Eye-Tracking Research

One interesting aspect of eye-tracking research is that even though we can detect fixations and say they are indicative of increased mental activity, we really do not know the reason the reader has fixated on that particular point. For example, it could be that s/he does not know the word in the first place, and so is trying to switch to a lower-level of processing in an effort to recompose it from its basic elements, or it could be that there is an error in the text that the mind has caught and is fixing in order to extract the information and integrate it into the overall whole. Specifically, eye-tracking research has been done in the area of word perception and identification. At this point it is only necessary to mention them as they relate to word recognition and phonological recoding.

In a study that looked at reading times of text with and without spaces between word boundaries, Everson (1986) concluded that beginning non-native readers of Chinese showed little or no influence due to word spacing, while advanced non-native readers seemed to be negatively influenced by the artificial spacing. In other words, the advanced non-natives took longer to read the passage than the beginning non-native readers, and their eye movement data indicated more fixations than those exhibited by the beginning non-natives. Further, native Chinese readers' performance did not suffer because of the artificial spacing, and Hayes (1990) echoes Everson (1986) when he suggests this was possibly because of "the robustness of their reading strategies" (Everson, 1986, p. 85; Hayes, 1990, p. 32).

Sun (1993) conducted an eye-movement study of native Chinese primary school, high school, and graduate students reading a text prepared in Chinese characters and Pinyin (a romanization system for Chinese). For comparison, the study also included two other groups: one consisted of presumably monolingual English speakers and the other of 10 Chinese-English bilinguals. The variety of reading materials included short paragraphs in both Chinese and English, short Chinese primary school texts written in both Chinese characters and Pinyin, individual characters, Chinese numerals. All five groups were shown individual Chinese characters.

Sun's (1993) data showed that without exception, the native readers fixated longer on the Pinyin, took in smaller and more frequent amounts of information during each fixation of the Pinyin text, and read the Pinyin text much slower than the character text (Sun, 1993, p. 252). Sun notes that "most of the books in China are printed in Hanzi (Chinese characters), so the greater difficulties associated with reading Pinyin may be due to a lack of practice" (p. 252) and concludes that, for the Chinese, "more efforts are required to master Pinyin" (p. 254).

Sun's (1993) study clearly showed the difference between three different readers of the Chinese character 靈 (líng, 'soul'), a character made up of 24 strokes and combining seven elements (雨 ['rain'], 口 ['mouth', x3], 工 ['labor'], 人 ['person', x2]), and divisible into two components (top and bottom—the phonetic霝 [líng] and the semantic 巫 [wú; 'wizard, witch']). Sun determined that degree of familiarity with a character guided eye movements. More importantly, based on his data, Sun concluded that:

> The reading eye movement data from bilingual subjects showed there are no real Chinese-English bilinguals. No one really possesses equal skill for reading Chinese and English quantitatively. For each bilingual subject, only one language is dominant, and the other has been suppressed. (Sun, 1993, p. 254)

3. RESEARCH QUESTIONS AND HYPOTHESES

The following two questions guided my research and will be addressed in my discussion of this study: 1) What does eye movement data tell us about cognitive load during computer adaptive tests (CATs)? 2) Are there identifiable differences between the three groups Non-Native Intermediate (NNI), Non-Native Superior (NNS), and Native Superior (NS) in terms of reading proficiency across orthographies?

The main hypothesis in this research is that there will be an increase in demands on cognitive load for test takers when working with a form of Chinese characters that is other than their declared preference. For example, for a learner who declares that s/he prefers the simplified Chinese characters, it is expected that when working with traditional Chinese characters they will exhibit more mental activity as evidenced by higher frequencies and durations of fixations.

There seems to be a prevailing trend among eye movement experts that these moments and locations are the instances when the brain is processing information. Key indicators of mental processing during reading are indicated by so-called eye fixations and saccades, moments when and locations where the eyes have paused or retrieved visual stimulus, respectively. What I looked for, then, are indications of "overload" in the comprehension process during reading of extended text by analyzing eye-tracking data of participants taking a CAT of Chinese reading proficiency.

3.1. Limitations of the Study

There is one significant limitation to the present line of inquiry in that the population of non-native learners of Chinese is much lower than that of more commonly taught languages such as French, German, and Spanish. In particular, Superior-level non-native readers of Chinese are extremely rare. A cursory examination of students currently enrolled in First- through Fourth-year Chinese language classes shows that by the fourth year of instruction, most students still do not achieve an Advanced level of reading proficiency. This affected the number of possible participants available for the Superior level, seriously limiting the quantity of data available for comparison and analysis, as well as the generalizability of the analyses.

4. METHOD

4.1. Participants

Data was gathered from a total of 25 participants (12 native English speakers [i.e., non-native Chinese] and 13 native Chinese speakers) divided into three groups: Group 1 consisted of 5 non-native students at the Intermediate level. Group 2 was made up of 7 non-native Chinese speakers at the Superior level. Group 3, the control group, consisted of 13 native readers. The Intermediate level participants had an average of two years of Chinese language study, while the non-native Superior level participants had an average of four years of formal language study. The native Chinese readers were of near-native proficiency or above in reading English. All participants were college-level readers (including undergraduate, graduate, and faculty) who had either normal or corrected-to-normal vision. Their ages ranged from 18 to 55 years. They were all paid US$15 for their participation. The data were gathered between March and August, 2003.

4.2. Materials

An ASL 501 head-mounted eye tracker was used to gather eye-tracking data. It measures a person's eye line of gaze with respect to their head. The eye tracker is actually part of a larger, complex system of components (for details, see Hayden, 2004).

4.2.1. Calibration Passages

Two reading passages were included in the calibration process at the beginning of each participant's session, the first in Chinese and the second in English. Each passage was two paragraphs in length and each paragraph filled one computer screen.

4.2.2. Computer-Adaptive (Proficiency) Test for Reading Chinese (CATRC)

The Computer-Adaptive (Proficiency) Test for Reading Chinese, or CATRC, is, as its name implies, a CAT that tests examinees' proficiency in reading Chinese. Altogether there are currently over 600 items in the CATRC, roughly 50 items per level (former ACTFL scale of nine levels from Novice-Low to Superior) with more items available at the Superior level. These items cover the four main proficiency levels (Novice, Intermediate, Advanced, and Superior), including all sublevels (i.e., Novice-Low, Novice-Mid, Novice-High, etc.), as described in the ACTFL Chinese Proficiency Guidelines (1987). The test randomly selects items to present to each examinee. When an examinee answers a question correctly, s/he is given an item of similar or greater difficulty; if s/he answers a question incorrectly the program then randomly selects an item of similar or lesser difficulty. Through this process the application identifies an examinee's reading level.

The CATRC has been shown to be reliable and valid for all levels ranging from Novice-Low to Superior. In particular, see Hayden (1998, 2003), Watanabe and Yao (1998), and Yao (1994, 1995). The CATRC fits into Bernhardt's (1999) idea of the essentials of comprehension assessment, in that it gives stable information about the abilities of learners so that their strengths and weaknesses can be identified; it ranks and orders comprehenders in their performance abilities; and, most important, it makes clear what they cannot *do* or *understand*" (p. 1, italics as in the original).

4.3. Procedures

The participants were all tested individually. They were seated in front of a Macintosh (G4, MacOS9) computer and were then shown the ASL 501 head-mounted eye tracker unit. Initial setup and calibration for each participant took between 5 and 15 minutes.

Calibration required the participants to look at an image of nine numbered points specifically arranged in a 3 by 3 grid that corresponded to the area of the visual field on the screen. After calibration participants were asked to orally read two short passages split into two paragraphs, each of which occupied one computer screen (1024x768 pixels) of text. The Chinese text was read first, and participants were able to choose between traditional or simplified Chinese character forms (depending on whichever form they felt more comfortable with). The English text was presented next.

After reading these passages, participants were calibrated again using the calibration image described above before beginning the CAT. The CATRC program first presents a screen with additional background questions mainly related to language study background. After the background information screen, there are three warm-up questions, then the actual test. Because the CATRC, as its name suggests, is a computer-adaptive test, participants varied widely on the length of time required to complete the assessment.

4.4. Data Collection and Analysis

Data collected from the eye tracker includes points of fixations, fixation durations, inter-fixation times, angular distance between fixations, and fixation sequences. Using the ASL software

(EYENAL and FIXPLOT) I was able to define areas of interest (AOI, collections of fixations in a given area), calculate gaze times (total fixation times per AOI), and identify saccades within the Chinese text as well as saccades between Chinese text, the English cue, and the English question and answers.

The CATRC keeps a record for each participant of the number of items tested, item ID (including level and item number), the correct answer for each item and the answer provided by the participant, a tally of correctly and incorrectly answered questions and each corresponding level, start and end times, a final reading proficiency level assessment based on the number of items answered correctly, and the total number of questions asked.

Data gathered from the CATRC provided descriptive statistics about each participant's performance on the test, including total test time duration, correct and incorrect responses, item difficulty levels, and assessed reading proficiency level. Data gathered from the eye-tracking system were analyzed for duration of eye fixations within each AOI, and elapsed time on task, as well as saccades across AOIs. The data for each participant was recorded into their own separate file.

Admittedly, the relatively small sample sizes of each group limits the generalizabilty of the findings; however, considering the population from which the sample came, there are highly suggestive trends that do make themselves apparent. As Everson (1986) noted almost twenty years ago, and is still very much the case with a language such as Chinese, researchers doing these types of studies "will probably have to resign him/herself to the inevitability of small sample sizes" (p. 91).

5. RESULTS

5.1. Quantitative Analyses

Table 1 shows that non-native Superior participants (NNS) required higher total fixations and higher mean fixations per item compared with Intermediate participants (NNI). This is most likely due to the fact that the NNS participants faced longer and more complicated CATRC items (including both passage and question) than the NNI participants. On the other hand, NNS participants also had higher mean fixations and higher mean fixation times per item compared with native Superior (NS) participants. This was expected, and would seem to indicate a point of language development of an emerging distinction between non-skilled and skilled readers (i.e., the NNS participants the NS participants, respectively).

Table 1. Group mean fixation data.

Group	*n*	Total Fixations	Mean Fixations/ Item	Mean Fixation Duration/Item (*ms*)
NNI	5	6738	190	338.00 (39.32)
NNS	7	8752	292	348.43 (89.12
NS	11*	4913	168	227.91 (29.70)

Note. Values in parentheses are SDs.
* One outlier (NS08) given 59 items due to a technical problem was removed from this group's calculations for this data set. This participant's information is otherwise listed in descriptive statistics throughout this paper. The data for native participant NS02 had to be eliminated due to excessive noise.

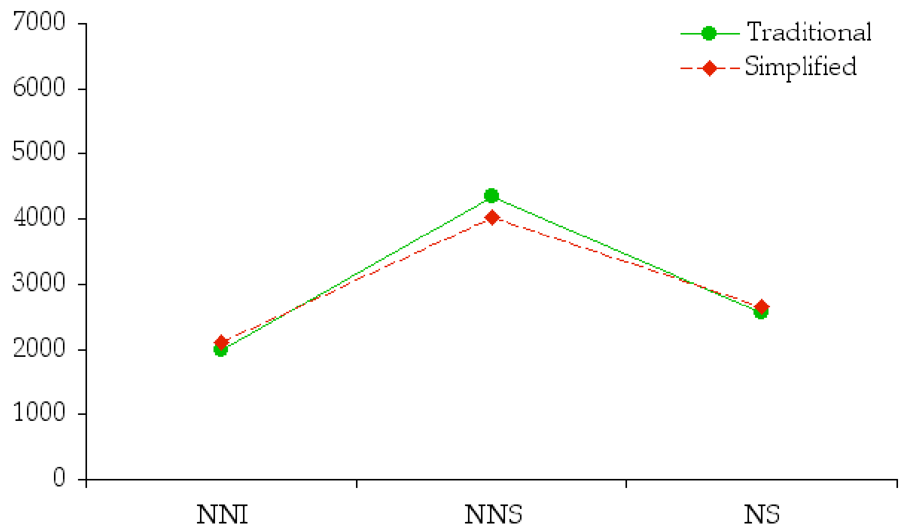

Figure 2. Fixation Frequency: Traditional vs. Simplified Characters.
Interaction between Chinese language ability and traditional/simplified characters as measured by fixation frequency.

Figure 2 shows similar performance for both the non-native groups (NNI and NNS) and the superior groups (NNS and NS). Although the mean fixation frequencies for the NNI group were about half as much as those for the NS group, mean fixation frequencies for traditional and simplified characters within each group were about the same. In other words, the mean frequencies for the NNI group and the mean frequencies for the NNS group each reading traditional Chinese characters were roughly equal to their group mean frequencies for reading simplified Chinese characters—meaning, perhaps, that both groups of readers found both forms equally difficult (or easy).

It is interesting that the NNS group is also similar to the NS in mean fixation frequencies for traditional and simplified characters. On one hand this suggests that both the NNS and NS participants dealt with both forms equally in the sense that neither form was more or less difficult to process per se. On the other hand, it also seems to suggest that the NNS participants had more difficulty in general with both forms of characters, although with a slight indication of strength for reading traditional Chinese characters (as could be expected from looking at their selected preference in Table 5).

5.2. Qualitative Analyses

The very small participant populations for each of the three groups in this study precluded the application of more complicated statistical measures of significance. The purpose of this section, then, is to actually look at example items to see if we can identify differences in reading based on the observed eye-tracking data. The key indicator for analyses will be fixation duration as represented by filled circles of varying sizes, i.e., smaller circles indicate shorter fixation durations while larger circles indicate longer fixation durations.

The best examples available for the analysis of this aspect of reading Chinese come from items at the Intermediate-Mid level and below. The majority of these items in the CATRC simultaneously present both traditional and simplified Chinese characters, generally with the former on top and the latter underneath in parentheses.

It is interesting to note that, depending on their preferred form of Chinese characters as selected at the beginning of the CATRC, NNI participants either stayed with or changed to the form they said they preferred. In addition to showing the influence of character complexity on fixation frequency and duration, Figure 3 also shows that a participant who has declared a preference for traditional Chinese characters will almost exclusively read that form when presented with a choice, and only rarely check the other form for clues to aid comprehension.

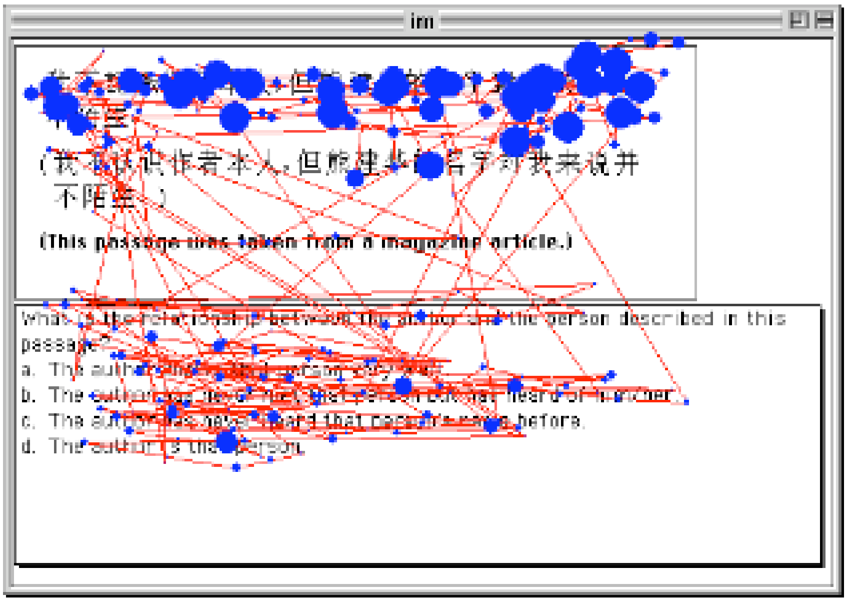

Figure 3. Eye-tracked Sample 2: NNI(T).
An example item (approx. 60% of original size) from a participant
who declared a preference for traditional Chinese characters.

Figure 4 shows how a NNI participant who declared a preference for simplified Chinese characters stayed almost exclusively with that form.

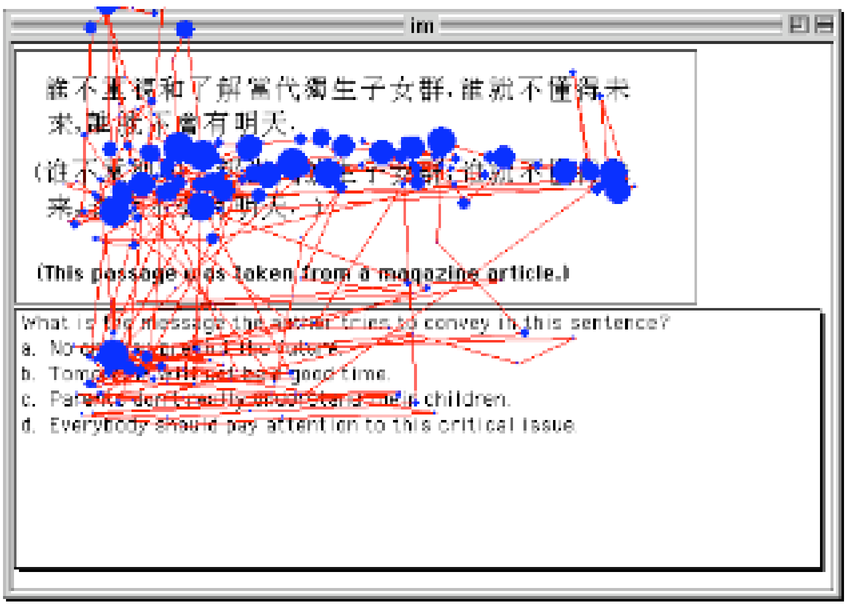

Figure 4. Eye-tracked Sample 2: NNI(S).
An example item (approx. 60% of original size) from a participant
who declared a preference for simplified Chinese characters.

Other examples show (not presented here in the interest of space; see the Appendixes in Hayden, 2004) participants changing mid-sentence from a non-preferred character form to a preferred character form. Instances such as this would have benefited the addition of a follow-up recall protocol to tease out from the participants more in-depth analyses about such observed reading behavior.

6. DISCUSSION

Looking at the fixation data provided in the previous section, we can see that NNI participants completed the CATRC with fewer fixations and in less time than the NNS participants. Also, for NNS participants, generally speaking, longer passages resulted in more fixations of longer durations. This, however, was not the case with NS participants encountering the same level, length, and difficulty of material, who generally required fewer fixations and less time to complete their assessment than NNS participants.

The most obvious reason for this becomes apparent through examination of the item set of each group upon completion of a test. Test items given to NNI participants were for the most part less complex (as measured in length of passage as determined by number of characters) and less difficult (in terms of content and subject matter) than those test items that the NNS and NS participants faced. Given the equal frequency of eye fixations among NNS participants during their time on task, we can infer that there was no increase in demand on cognitive processes for non-preferred character form.

The eye tracking data shows us that, despite the relatively shorter passage content, non-native Intermediate readers (NNIs) of CFL spend long spans of time working through the test items. This suggests that NNIs are generally concerned with processing each and every character for meaning, if not sound. Failure to come up with either meaning or sound or both may cause a bottleneck in comprehension processes delaying the “click of comprehension” that signals to the reader that they have comprehended something.

Everson (1986) posits that “class time can be devoted to exercises designed to encourage as well as enhance the learner’s silent reading skills” (p. 90, emphasis mine). In contrast, Dew (1994) notes that “reading and the expansion of vocabulary is for the most part a solitary activity, undertaken by the student in his own unsupervised time" (p. 40). It may be that both of these points of view coexist on one continuum of reading instruction whereby during initial stages of language instruction students receive explicit guidance and practice in how to effectively navigate the reading process, and moving towards autonomy and independence only occasionally needing to consult common resources such as dictionaries and online search engines.

From the data presented in the present study, the reading strategies of non-native Superior readers generally differ from that of non-native Intermediate readers, but are still not as efficient as those of native Chinese readers. So, we may well ask the following question: "How can learners at lower levels of CL2/CFL reading proficiency be best taught to approach novel (i.e., unfamiliar) reading tasks, based on the processing strategies exhibited by native readers?" It would seem that a wide variety of graded readers needs to be collected and/or created in order to allow novice-, intermediate-, and even advanced-level CFL students more opportunities to gain more exposure to the written language. The greatest hindrance to incorporation of extensive reading in programs for languages such as Chinese and Japanese is the general dearth of materials available that will actually turn emerging readers into avid readers.

Since the CATRC does not include items at the Distinguished level of the ACTFL reading proficiency guidelines, this suggests either a limitation in the precision of the CATRC or of the ACTFL reading guidelines themselves, as suggested by Park (1999). Future research in this area might need to develop a more precise assessment instrument that has a large item bank (perhaps at least 100 items) for each level of proficiency and that includes items at the Distinguished proficiency level. However, as Horodeck (1987) notes, tactics such as "skimming, skipping around, and guessing" are what real readers do. So, in that sense, the ideal assessment instrument would allow for participants to use the same skills as they would in "real reading", as opposed to a test that was designed differently (e.g., to see whether readers could spot errors where their reading strategies would mainly be those characteristic of proofreading). Advances in eye-tracking equipment and technology now allow researchers to gather data across multi-screen and multi-page scenes, which would mean that studies could be done that look at non-native and native readers of Chinese reading lengthy paragraphs, pages, chapters, and whole books.

7. CONCLUSION

As with the Everson (1986) study, the purpose of this study was multifaceted: first, to add to the still paltry collection of research that looks at non-native readers of Chinese. While many studies have recently added much to the eye-tracking research on native readers of Chinese, there is still much work to be done. Second, in this study I have attempted to highlight the phenomenon of cognitive load that is involved during the reading of novel text as presented via a computer-adaptive test (CAT). By using a CAT, it was possible to expose readers to both orthographies (i.e., English and Chinese), as well as two forms of a single orthography (i.e, traditional and simplified Chinese characters), requiring them to vary their reading strategies during the administration of a single bilingual reading proficiency test.

In concord with Everson's (1986) conclusions, the data gathered and analyzed in this paper supports perceptual models of reading which takes efficiency in the perception of the printed word as a measure of reading development (p. 81). Based on the reading performances of non-native Intermediate (NNI) and non-native Superior (NNS) level readers in comparison with native readers of Chinese, we can conclude that the NNI readers still do not have the familiarity or experience with the written Chinese language necessary to allow for automatic transitioning between perceptual subskills that would make for more efficient reading comprehension. Similarly, as with Everson's Advanced readers, the NNS readers in the present study also generally do not have a level of automaticity needed for efficient reading at a native level as exhibited by the NS readers in this study.

REFERENCES

ACTFL Chinese Proficiency Guidelines. (1987). *Foreign Language Annals, 20*, 471-487.

American Council on the Teaching of Foreign Languages. (1986). *ACTFL proficiency guidelines*. New York: American Council on the Teaching of Foreign Languages.

Bernhardt, E. B. (1999). If reading is reader-based, can there be a computer-adaptive test of reading? In M. Chalhoub-Deville (Ed.), *Studies in language testing 10: Issues in computer-adaptive testing of reading proficiency* (pp. 1-10). Cambridge: University of Cambridge Press.

Chen, H-C. (1986). Component detection in reading Chinese characters. In H. S. R. Kao & R. Hoosain (Eds.), *Linguistics, psychology, and the Chinese language* (pp. 1-10). Hong Kong: Centre of Asian Studies.

Chen, H-C. (Ed.). (1997). *Cognitive processing of Chinese and related Asian languages*. Hong Kong: Chinese University Press.

Chen, H-C., & Tzeng, O. J. L. (Eds.). (1992). *Advances in psychology 90: Language processing in Chinese*. New York: North-Holland.

DeFrancis, J. (1966). Why Johnny can't read Chinese. *Journal of the Chinese Language Teachers Association, 1*(1), 1-20.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems*. Honolulu, HI: University of Hawaii Press.

Dew, J. E. (1994). Back to the basics: Let's not lose sight of what's really important. *Journal of the Chinese Language Teachers Association, 29*(2), 31-46.

Everson, M. E. (1986). *The effect of word-unit spacing upon the reading strategies of native and non-native readers of Chinese: An eye-tracking study*. Unpublished doctoral dissertation, The Ohio State University.

Everson, M. E. (1994). Toward a process view of teaching reading in the second language Chinese curriculum. *Theory into Practice, 33*(1), 481-507.

Feldman, L. B., & Siok, W. W. T. (1999). Semantic radicals in phonetic compounds: Implications for visual character recognition in Chinese. In J. Wang, A. W. Inhoff, & H-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum Associates.

Guder-Manitius, A. (1999). *Sinographemdidaktik: Aspekte einer.systematischen Vermittlung der chinesischen Schrift im Unterricht Chinesisch als Fremdsprache*. Heidelberg: Julius Groos Verlag.

Hayden, J. J. (1998). The influence of a semester abroad on reading proficiency: A descriptive study. *Journal of the Chinese Language Teachers Association 33*(3), 13-24.

Hayden, J. J. (2003). Shocking our students to the next level: Language loss and some implications for teaching Chinese as a foreign language. *Journal of the Chinese Language Teachers Association 38*(3), 1-20.

Hayden, J. J. (2004). *Why Johnny* can *read Chinese: Working memory, cognitive processes, and reading comprehension*. Unpublished doctoral dissertation, University of Hawaiʻi at Mānoa.

Hayes, E. B. (1987). The relationship between character complexity and character recognition. *Journal of the Chinese Language Teachers Association, 22*(2), 45-58.

Hayes, E. B. (1990). The relationship between 'word length' and memorability among non-native readers of Chinese Mandarin. *Journal of the Chinese Language Teachers Association, 25*, 31-41.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Horodeck, R. A. (1987). *The role of sound in reading and writing* kanji. Unpublished doctoral dissertation, Cornell University.

Hsueh-chao, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403-430.

Ke, C. (1996). An empirical study on the relationship between Chinese and character recognition and production. *Modern Language Journal, 80*, 340-349.

Matsunaga, S. (1994). *The linguistic and psycholinguistic nature of* Kanji*: Do* Kanji *represent and trigger only meanings?* Unpublished doctoral dissertation, University of Hawaiʻi at Mānoa.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

Moser, D. (1991). Why Chinese is so damn hard. In V. H. Mair (Ed.), *Schriftfestschrift: Essays in Honor of John DeFrancis, Sino-Platonic Papers, 27* (pp. 59-70). Philadelphia, PA: University of Pennsylvania.

Park, S. (1999). *Testing the EFL skills and text hierarchy of the ACTFL reading guidelines*. Unpublished master's thesis, Honolulu, University of Hawaiʻi at Mänoa.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(1), 101-118.

Sergent, W. K., and Everson, M. E. (1992). The effects of frequency and density on character recognition speed and accuracy by elementary and advanced L2 readers of Chinese. *Journal of the Chinese Language Teachers Association 27*(1/2), 29-44.

Sun, F. (1993). Eye movements in reading Chinese: Paragraphs, single paragraphs, and Pinyin. In S. F. Wright & R. Groner (Eds.), *Facets of Dyslexia and its Remediation* (pp. 245-256). Amsterdam: North-Holland.

Tsai, C-h. (2001). *Word identification and eye movements in reading Chinese: A modeling approach*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign.

Tzeng, O. J. L., Hung, D. L., & Garro, L. (1978). Reading the Chinese characters: An information processing view. *Journal of Chinese Linguistics, 6*, 287-305.

Wang, M., Perfetti, C. A., Liu, Y. (2003). Alphabetic readers quickly acquire orthographic structure in learning to read Chinese. *Scientific Studies in Reading, 7*(2), 183-208.

Wang, A., & Thomas, M. (1992). The effect of imagery-based mnemonics on the long-term retention of Chinese characters. *Language Learning 42*(3), 359-376.

Watanabe, Y., & Yao, T-c. (1998). Examining the validity of the Computer-Adaptive Test for Reading Chinese. Paper presented at the Second Annual College-wide Conference for Students in Languages, Linguistics, and Literatures, University of Hawaiʻi, Honolulu, HI, March 1998.

Xiao, Y. (2002). The effect of character density on learning Chinese as a foreign language. *Journal of the Chinese Language Teachers Association 37*(3), 71-83.

Yang, H-M., & McConkie, G. W. (1999). Reading Chinese: Some basic eye-movement characteristics. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 207-222). Mahwah, NJ: Lawrence Erlbaum Associates.

Yao, T-c. (1994). Computer-Adaptive Test for Reading Chinese [Computer software]. Honolulu: University of Hawaiʻi at Mānoa.

Yao, T-c. (1995). A computer-adaptive test for reading Chinese (CATRC): A preliminary report. *Journal of the Chinese Language Teachers Association 30*(1), 1-20.

Zhou, X., Marslen-Wilson, W., Taft, W., & Shu, H. (1999). Morphology, orthography, and phonology in reading Chinese compound words. *Language and Cognitive Processes, 14*(5/6), 525-565.