Can Word Segmentation and Length of Text Affect L2 Reading? Evidence From Advanced L2 Learners of Chinese and the Three Reading Models

Ken Chen
Shanghai International Studies University (SHISU), Shanghai, China

Abstract—**The purpose of this article is to examine whether the word segmentation and the length of text will promote Chinese as a second language reading, and what the appropriate reading models are best for L2 learners when they processing Chinese language reading tasks. According to the experiment test, the results indicated that the main effects of word segmentation and length of text were not significantly different, and the interaction effect of the two factors was not significant in statistics too for advanced proficiency L2 learners of Chinese. This article concludes by using the approach of word boundary and choosing the useful reading models for teaching Chinese reading in classrooms are the valuable approaches for teachers when they combine other effective ways in L2 education.**

Index Terms—**word segmentation, length of text, reading models, advanced L2 learners of Chinese**

I. INTRODUCTION

For most readers who use phonetic characters, there are always physical boundaries between words in their written texts, so there is no confusion between adjacent words. However, in the ideographic writing system with Chinese as the typical representative, there is no word segmentation, and also no natural boundary between words, but only the physical boundary between every character. Zheng (1981) asked the participants to identify the position of any Chinese character in the fast presentation of Chinese characters from the meaningful words (real words) composed of two Chinese characters and the meaningless words (non words) composed of the two Chinese characters. The results showed that the accuracy of recognition of native Chinese readers under the condition of real words was significantly higher than that under the condition of non words. Does the presentation of written Chinese text have an impact on the reading of Chinese as a second language? The research on this question is helpful to understand some deep mechanism of Chinese reading.

Gao (2006) conducted a survey on foreign students from Japan, South Korea and other countries with Chinese character culture circle, and found that word segmentation did not have any impact on such students' Chinese reading. Zhou and Xie (2007) conducted sentence and word segmentation experiments on 38 intermediate proficiency Chinese as a second language learners from Chinese character culture circle through oral test and paper-pencil test. Through the analysis of the experimental data, they pointed out that the reasons for the errors in word segmentation and put forward some strategies for word segmentation. Wang Di (2011) used spaces and shadows as word segmentation markers, and used DMDX program to record the reaction of Russian students who studied Chinese as a foreign language at different levels when reading Chinese declarative sentences. The results revealed that the Russian students had the shorter reading times under the condition of word segmentation. At the same time, their reading times under the condition of shadow word segmentation were relatively short. Word segmentation can promote the Russian students' Chinese reading speed, and students at high level spend less reading times than those who at middle and low level. Song (2014) conducted a text segmentation experiment on Indonesian and Thai L2 learners of Chinese. The results indicated that the higher the learners' Chinese proficiency, the stronger their word segmentation ability and the reading units will expand with the improvement of Chinese proficiency. Song (2014) conducted an experimental study on the influence of word segmentation on Chinese reading for foreign students. The experimental results showed that the primary level students were more suitable for the Chinese articles in the form of word segmentation, which had a certain role in promoting their reading. However, word segmentation had no obvious effect on the Chinese reading of intermediate and advanced students. Gao and Jiang (2015) conducted word segmentation and Chinese reading tests on Chinese second language learners with Pinyin background by means of paper-pencil test. The content of the experiment was the comprehension of different length articles with and without word segmentation. The results revealed that the effect of word segmentation on the reading ability of the participants was positively related to the length of the text. In addition, the research also revealed that the processing patterns of Chinese reading of primary and intermediate proficiency L2 learners of Chinese were very similar, which indicated that primary and intermediate levels of learners were likely to

rely mainly on the bottom-up Chinese reading comprehension pattern.

To explore whether the presentation of word segmentation and the length of a text will affect reading speed and accuracy of L2 learners of Chinese, there are two research questions this article focuses on:

1. Does word segmentation in text promote and decrease L2 learners' reading times?
2. Does text length affect L2 readers' reading comprehension?

II. METHODS

A. Participants

This study selected 20 advanced L2 learners of Chinese who studied in China's universities as participants, including 9 males and 11 females. Their Chinese levels reached the advanced stage, with an average age of 25.35 and an average Chinese language learning time of 4.73 years. The visual acuity or corrected visual acuity of all participants was normal, and there were no hearing impairment and dyslexia participants.

B. Materials

The tested materials used in the experiment were texts, and the selection process was as follows:

First of all, several learners who did not participate in this experiment selected the content areas they were interested in from the "*People's Daily* newspaper page names and corresponding contents / fields (categories) list" sorted out by the researcher. The researcher calculated statistics of the top ten content areas.

Secondly, the researcher selected the corresponding texts from the selected content areas, and only one text was selected as the experimental materials for each content area. The texts in these content areas were derived from the newspaper corpus of the *People's Daily* with an event span of one month. The selected experimental materials fully reflected event content and involved all kinds of news topics and language styles.

Thirdly, the learners who did not participate in this experiment read the experimental materials selected by the researcher to determine the acceptability and readability of each text. If a certain text was considered by most learners to be more difficult to accept or less readable, the researcher would reselect the materials in the corresponding content area until all the experimental materials had passed the acceptance and readability tests.

Fourthly, the number of words (including punctuation) of ten selected materials texts was calculated to count the length of the corpus. The average number of words was 197.70. The text less than or equal to the mean value was short, otherwise it was long. In this way, five of the ten texts were short and five were long. After the length was determined, all the learners who did not participate in this experiment were asked to make an acceptable judgment on the division of the length. After the judgment of acceptability, all the texts were segmented word by word by the researcher.

Fifthly, referring to the length and word boundary conditions, 10 participants used the non-word segmentation of "3 short texts + 2 long texts" and the word segmentation of "3 long texts + 2 short texts", while the other 10 participants used the non-word segmentation of "3 long texts + 2 short texts" and the word segmentation of "3 short texts + 2 long texts". In this way, each participant would read 10 texts, and five texts under two word boundary conditions and two types of length conditions.

C. Procedures

The participants were asked to complete the reading comprehension multiple-choice questions immediately after reading a text. Each experimental text was printed on 1/2 of A4 paper, one side was the material, and the other side was the reading question which needed the participants to choose the correct answer. The question of each text was two single-choice questions. The participants could operate according to the "instruction" of the experiment. There was no limit to the reading time, but when the participants finished the two questions, they could not look back at the experimental texts.

The concrete procedures of this test were as follows:

First of all, the researcher explained the intention of this experiment, the process of this experiment operation and the related precautions to the participants, and answered the questions about the experiment operation.

Secondly, after the participants understood the operation process of this experiment, they began the experiment. During the experiment, the participants were not allowed to consult dictionaries, textbooks and other materials, and they completed all the reading comprehension tasks independently.

Thirdly, after the participants completed all the experimental tests, the researchers immediately checked the participants' answers to ensure that the data were complete and correct.

The experiment was conducted in a separate way, and the relevant instructions before the experiment were in Chinese.

III. RESULTS

This experiment used the two factors within-group design to calculate the related statistic data. The independent variables in this experiment were word segmentation and text length. Word segmentation was divided into two levels, non-word segmentation and word segmentation. Text length was also divided into two levels, short text and long text.

The dependent variable was the number of reading comprehension errors.

The participants completed all the single-choice questions of reading comprehension. One wrong choice was counted as 1 point and two wrong choices were counted as 2 points, and so on. No wrong choice was counted as 0 point. The cumulative number of errors was only calculated in the same category of experimental data. Spss22.0 was used to calculate the experimental data.

TABLE 1
THE DESCRIPTIVE STATISTICS OF ERROR NUMBERS IN READING COMPREHENSION（N = 20）

	non-word segmentation				word segmentation			
	short text		long text		short text		long text	
	M	SD	M	SD	M	SD	M	SD
Error numbers	0.85	0.88	0.60	0.82	1.20	0.89	0.65	0.68

The data in Table 1 were tested by analysis of variance (ANOVA), and the results are as follows:

(1) The main effect of word segmentation on the number of errors in reading comprehension was not significant (f (1, 19) = 1.652, P = 0.214 > 0.05). There was no significant difference in the number of reading comprehension errors between the texts with word segmentation and those without word segmentation (MD = 0.35) under short texts condition. There was no significant difference in the number of reading comprehension errors between the texts with word segmentation and the texts without word segmentation under long texts condition (MD = 0.05).

(2) The main effect of length on the number of reading comprehension errors was not significant (f (1, 19) = 3.848, P = 0.065 > 0.05). Under the condition of no word segmentation, there was no significant difference in the number of reading comprehension errors between short length and long length (MD = 0.25). Under the condition of word segmentation, there was no significant difference in the number of reading comprehension errors between short length and long length (MD = 0.55).

(3) The interaction between word segmentation and text length was not significant (f (1, 19) = 1.132, P = 0.301 > 0.05), and there was no significant difference between word segmentation and text length on the number of reading comprehension errors. The experimental interaction was not significant, which indicated that word segmentation and text length had the same effects on the participants' reading.

IV. DISCUSSION

A. Word Segmentation and Text Length on Reading

The results revealed that word segmentation and text length both had no effect on the number of errors in reading comprehension, and the interaction between them had no effect on the text reading. Further research showed that under the word boundary condition, the number of errors in reading comprehension of short and long texts was slightly less than that of texts with word segmentation. It could be seen that the word boundary information had little influence on the advanced stage participants when they read the text. On the other hand, under the same word segmentation condition, the number of reading comprehension errors in short texts was slightly more than that in long texts. From this result, it could be seen that the higher-level participants' comprehension and information integration of the longer discourse were better than those of the phrase discourse and the longer discourse provides more contextual information and language content.

Additionally, the results of Gao and Jiang (2015) text reading experiments on foreign students at the primary and intermediate stages also showed that the main effect of word boundary conditions on the number of reading errors was not significant. However, in their experimental results, the interaction of text length and article word boundary condition had significant effect on the number of reading comprehension errors. Through comparative observation and analysis, there were several differences between the experimental materials of this article and that of Gao and Jiang's.

First, the source of the tested texts was different. In this experiment, the researcher used the existing, unadapted, and native Chinese language oriented newspaper corpus. However, the materials used by Gao and Jiang were specially designed for the participants. Their tested materials were the Chinese reading materials compiled by researchers according to the existing Chinese reading and listening textbooks.

Second, the experimental design was different too. After reading the tested materials, the participants were asked to complete a closed reading comprehension test (single-choice test). But in the Gao and Jiang's experiment, the participants were asked to complete an open reading comprehension task (using Chinese characters or pinyin to answer subjective questions) after reading the tested materials.

The differences of tested materials' sources and questions design may lead to differences in the results of the text reading experiment. However, by comparing the two groups of discourse experiments, we discovered that the word boundary condition had no effect on the Chinese texts reading of the foreign students at the three stages of elementary, intermediate, and advanced levels, while the text length had a certain effect on the foreign students who at the elementary and intermediate levels. The result of this comparison proved once again that advanced L2 learners of Chinese rely more on top-down reading strategies in text reading.

B. Reading Models When Processing L2 Texts

Text reading is a complex and multidimensional psychological process of linguistic information. This process involves many steps of linguistic information processing at different levels, from the recognition of words, the understanding of sentence meaning to the construction of text meaning. Text reading comprehension is not only the process of the reader's immediate information, but also the integration of readers own existing background knowledge and language knowledge. Through the reorganization of the short-term processed language information and the long-term stored language knowledge, the readers can obtain new language content. There are three types of text reading models which includes bottom-up model, top-down model, and interactive model.

According to the bottom-up model of text reading, the reading process is an orderly bottom-up cognitive process from word to word, sentence to text. If the reader wants to extract and understand the text information well, he or she must rely on the understanding of the meaning of the sentences that compose the text. To understand the meaning of a sentence, readers must rely on the recognition and meaning understanding of the words that directly constitute the sentence, and the understanding of the words has to go to the recognition and meaning understanding of the Chinese characters that constitute the words. The reading process under this model is an extremely accurate cognitive processing process. The reader is gradually transiting from the perception, identification, processing and understanding of language symbols at the bottom to the perception, identification, processing and understanding of language symbols at the top. Therefore, as far as reading comprehension is concerned, it is a cognitive process of separation and integration of the language and words that constitute the text.

The bottom-up reading model is more useful for analyzing the primary stage of second language learners, because they are relatively lack of target language knowledge and are used to seeking language information from a single component. However, the intermediate and advanced stage of second language learners have stored appropriate target language knowledge in their mind, and they may be more affected by their own existing language knowledge and background knowledge in the process of text reading. For example, a certain letter is easier to be recognized in meaningful words than in meaningless and arbitrarily arranged strings, which is called "word optimization effect", which shows the influence of readers' existing language knowledge on their reading.

According to the top-down model of text reading, it is not a cognitive process from word to word, sentence to text, but a process of selecting, identifying, recognizing and processing the present text information based on the reader's existing knowledge of the objective world and language. Therefore, the extraction and understanding of text information is a top-down process based on the reader's prior knowledge and concept driven. For example, a word is easier to recognize in a context of definite meaning than when it appears alone, which is called "contextual effect".

The top-down reading model also has some problems. It overemphasizes the prediction of text information through the reader's existing background knowledge and language knowledge, and ignores the identification of the necessary lexical and syntactic structures. Some studies have shown that it is not the predictive ability of language information that makes the second language learners with stronger reading ability, but more importantly their decoding ability of the target text, which makes them have stronger reading ability. Second language learners' text reading ability is always from low to high. The foundation of their fast and accurate text information extraction and understanding is still their language information decoding ability.

Both bottom-up and top-down text reading models follow the principle of sequence processing of linguistic information. Information processing process is independent, and the previous processing results will be transferred to the subsequent processing stages, without mutual influence. The difference between them is that the former indicates that the reader has experienced the process of text cognitive processing from low level to high level, while the latter reveals that the reader's text reading is to verify the hypothesis and prediction of the text before reading after they get the language information. Each of the two reading models has its own limitations. The bottom-up reading model ignores the reader's existing language knowledge and overemphasizes the role of the reader's decoding ability in text information extraction and comprehension. On the other hand, the top-down reading model pays one-sided attention to readers' experience and ignores their decoding ability in the face of specific discourse content.

Reading is a complex psychological process. Advanced stage learners do not always read Chinese characters one by one, but may complete the reading in the form of phrases or chunks. The learners in this stage basically have a relatively complete ability of Chinese words, and they are closer to the reading habits and ability of native Chinese learners than the learners in the early and middle levels. The results of this experiment showed that word segmentation had no significant effect on the reading process, regardless of the length of the text. This revealed that the advanced Chinese learners no longer simply rely on the meaning of a single word to infer the meaning of the whole text. They may use a larger combination of language units to promote the reading comprehension of the text. Advanced stage learners have more knowledge of Chinese words and corresponding cultural knowledge, but also have better Chinese reading ability, and their perceptual range of reading is significantly larger than the coverage of a single Chinese word. It is undeniable that although the Chinese reading skills of advanced stage learners tend to be similar to those of native Chinese learners, it is difficult for them to achieve the reading skills of native Chinese learners in the end. When they encounter more difficult words in the process of reading, they may also take a more primitive way, that is, a single word processing way, to read. Therefore, there is a space of proportional distribution among learners of different levels in the three reading models of "bottom-up", "top-down" and "interactive reading style". For advanced stage learners, they may use the

"top-down" reading method in the reading process.

With the development of text reading experiment, more and more evidence shows that when reading the target text, the second language does not simply use one text reading model, but uses the bottom-up and top-down text reading models. When it comes to specific text reading, it is the question of how much of the two models are used. The interactive model of text reading has an impact on the cognitive processing stages of words, syntax, text and semantics. Schema theory is a typical representative of this interactive reading model.

V. CONCLUSIONS

This research conducted an empirical study on the texts reading of advanced L2 learners of Chinese. The results indicated that both word segmentation and text length had no significant effect on reading. Through the experiment, it was shown that the advanced Chinese as the second language learners may use more "top-down" reading style in the reading process. At this stage, the learners have acquired good Chinese words ability and relatively complete reading skills. In this stage of Chinese reading course, we can consider the following teaching suggestions.

First, increase the length of reading texts, integrate more word knowledge and cultural knowledge, and try to choose real-life texts without any manual modification, such as newspaper texts, academic texts and novel texts, so as to promote learners' access to more authentic language materials and improve their ability to solve practical communication problems in the target language.

Second, the reading materials can be articles with more chunks, which not only conforms to the language reading level of advanced learners, but also can continuously improve their reading skills and effectively promote their language information processing ability.

Third, it is necessary to increase the difficulty of words in the reading materials and select a certain number of super ordinate words, so that learners can solve the problems of words in the reading process and fully mobilize their subjective initiative through contextual information and semantic relevance of related words.

Generally speaking, in the actual teaching of text reading, teachers should give full consideration to the target language knowledge that the second language has mastered before, so as to ensure that the learners have enough language ability to complete the learning of target language. On the other hand, teachers should also consider the difficulty of the text and the adaptability of the learners' knowledge and ability of the target language. Appropriately, teachers can consider choosing a text that is slightly higher than learners' current language ability for teaching.

REFERENCES

[1] Bai, X., Yan, G. & Liversedge, S. P. et al. (2008). Reading Spaced and Unspaced Chinese Text: Evidence from Eye Movements. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1277-1287.
[2] Bai Xuejun. & Tian J. et al. (2009). An Eye-tracking Research on Chinese Reading under Word Segmentation Condition by American College Students. *Journal of Nankai Linguistics*, 1, 140-153.
[3] Gao Shan. (2006). The Effect of Word Boundary Information on Chinese Reading with Overseas Students (Ci Bianjie Xinxi dui Liuxuesheng Hanyu Yuedu de Yingxiang). Master Thesis, Beijing Language and Culture University.
[4] Gao Shan. & Jiang X. (2015). The Effect of Word Boundary on Chinese Reading with L2 Learners (Ci Bianjie dui Dier Yuyan Xuexizhe Hanyu Yuedu de Yingxiang). *Language Teaching and Linguistics Studies*, 4, 8-16.
[5] Li, X., Rayner, K. & Cave, K. P. (2009). On the Segmentation of Chinese Words During Reading. *Cognitive Psychology*, 58: 525-552.
[6] Song Kailun. (2014). The Effects of Word Boundary Information on Reading to Different Levels of Indonesian and Thai L2 Learners of Chinese (Ci Bianjie Xinxi dui Butong Hanyu Shuiping Yinni he Taiguo Liuxuesheng Yuedu de Yingxiang). Master Thesis, Jinan University.
[7] Song Yue. (2014). The Effect of Word Segmentation on Reading to Overseas Students in China (Ci Qiefen dui Laihua Liuxuesheng Hanyu Yuedu de Yingxiang). Master Thesis, Nanchang University.
[8] Wang Di. (2011). The Effect of Word Segmentation on Chinese Reading to Russian Students (Ci Qiefen dui Eluosi Xuesheng Hanyu Yuedu de Yingxiang). Master Thesis, Tianjin Normal University.
[9] Zhou Jian. & Xie H. (2007). The Study on Chinese Reading Ability of Word Segmentation and Semantic Retrieving to Overseas Students (Liuxuesheng Hanyu Yuedu Fenci he Yuyi Tiqu Nengli Yanjiu). *Journal of Chinese Learning*, 2, 70-76.
[10] Zheng Zhaoming. (1981). Perception of Chinese Characters. *Acta Psychologica Taiwanica*, 23, 12-19.

Ken Chen was born in Chengdu, China in 1990. He is a Ph.D. student at the School of Chinese Studies and Exchange, Shanghai International Studies University, China. His research interests include second language acquisition, L2 cognitive processing and Chinese for specific purposes.