The Relative Significance of Vocabulary Breadth and Syntactic Knowledge in the Prediction of Reading Comprehension Test Performance

SHEN Yalin
Hefei Normal University

TAO Wei
University of Science and Technology of China

Abstract

Founded on the componential model of reading, Nation’s construct of vocabulary breadth test and structuralists’ view of syntactic knowledge, this paper attempts to investigate the relative significance of vocabulary breadth and syntactic knowledge in the prediction of Chinese EFL learners’ reading comprehension test performance and whether the relative significance is moderated by the learners’ different L2 proficiency. An experiment including three tests was carried out with 68 sophomores in Anhui Medical University. The findings show that: 1) there is a positive linear correlation between the learners’ performance on a reading comprehension test and their vocabulary breadth and syntactic knowledge, with the multiple r being 0.551; 2) syntactic knowledge outperforms vocabulary breadth in predictive power and emerges as the stronger predictor of reading comprehension test performance; and the relative significance of syntactic knowledge over vocabulary breadth in the prediction of reading comprehension does not change with the EFL learners’ L2 proficiency. The findings of this study provide implications for both reading instruction and remediation of learners’ reading problems.

Key words: vocabulary breadth; syntactic knowledge; L2 reading comprehension

1. Introduction

Among the different approaches to research in reading, the componential approach is concerned with identifying possible explanatory components or skill factors involved in
The Relative Significance of Vocabulary Breadth and Syntactic Knowledge in the Prediction of...

the reading process, rather than explaining how those components operate in the process. Carr and Levy (1990: xi) see value in such an approach: “Many investigators believe that the kind of full characterization that results from component skills analysis is the only way to get an accurate picture of reading ability, how it changes developmentally, and what creates individual differences among readers…” In the componential approach to modeling reading ability, a number of contributory factors have been empirically validated. But research on their relative contribution to explaining performance on second language reading tests is limited. Furthermore, the contribution of syntactic knowledge has been largely ignored in comparison with the attention to focused on vocabulary. However, consideration needs to be given to the factor of syntactic knowledge in such comparisons in adult L2 reading. Early support for this is found in such publications as Berman (1984), and later in Urquhart and Weir (1998: 269), where they note:

Grammar is a component of reading that has been almost ignored in the research. It seems to us that this is an interesting and potentially valuable research area which L2 teachers and applied linguists are in a good position to investigate.

The limited literature on the significance of syntactic knowledge for L2 reading is inconclusive (Ulijn, 1981, 1984; Alderson, 1993; Bernhardt, 1999) and the literature on the relative contribution of the vocabulary and the syntactic knowledge to reading performance is too limited to offer convincing evidence for supporting one or the other of the two predictors (Bernett, 1986; Bossers, 1992; Brisbois, 1995; Haynes & Carr, 1990; Toshihiko & Weir, 2007; Van et al., 2004). Hence the limited literature on the relative significance of the two components of reading ability would benefit from further empirical research, preferably based on samples which are larger than and different from most of the previous studies. The present research aims to investigate the relative contribution of Chinese college students’ L2 vocabulary breadth and L2 syntactic knowledge to L2 reading comprehension. To be specific, the present study intends to address the following two questions:

1) To what extent do L2 vocabulary breadth and L2 syntactic knowledge predict reading comprehension performance? Which has stronger predicting power?
2) Does L2 proficiency affect the predicting power of L2 vocabulary breadth and L2 syntactic knowledge?

2. Theoretical Framework

2.1 Componential Model of Reading
The componential model of reading is developed to explain reading ability. To be specific, it is to “identify and isolate separate factors that constitute the ability of reading and try to operationalize these components into skills and strategies” (Jin, 2002: 84). Compared with the process model of reading which describes the reading process sequentially according to the temporal order, the componential reading model avoids dealing with the abstract
and invisible reading process but investigates specific reading skills underlying the process. In addition, through the separation and identification of components of reading skills, the model provides an operational definition of reading ability, which proves to be facilitative of improving reading instruction and developing reading tests and would further contribute to the construct validity of the test.

According to the number of componential skills isolated, Rost (1993) has summarized three types of componential models: general-factor theories, the two-factor model and the multiple-factor model. Compared with the first two types, the multiple-factor model, which views reading ability as being made up of more than three components, is a useful approach, because it offers important insights into the reading process (Grabe, 1991; Lunzer et al., 1979).

2.2 Definition and Assessment of Vocabulary Breadth
One dimension of vocabulary knowledge as defined by Anderson and Freebody (1981) is called “breadth” of knowledge, which means the number of words for which a person knows at least some aspects of meaning.

Schimitt (1999) states that, in the field of L2 lexis, there are no accepted standardized tests available to assess vocabulary breadth. The closest things to standardized tests are the Levels Test (Nation, 1990) and a number of checklist tests developed by Meara and his associates (Meara & Buxton, 1987; Meara & Jones, 1990).

The Vocabulary Levels Test uses the definition-matching technique to test the knowledge of words of different frequency levels. This is a pen-and-paper test which includes five parts representing five frequency levels, that is, the first 2000 words, 3000 words, 5000 words, the University word level and 10,000 words. These five levels are primarily defined by reference to the word-frequency data in Thorndike and Lorge (1994). For each frequency level, the test consists of a number of item clusters, and the test takers are required to match half of the words to short definitions of their meanings. In the Levels Test, the words are presented in isolation. Read (1997) points out that in order to get a reliable estimate of vocabulary size, the sample of the words tested needs to be quite substantial. That is to say, a relatively decontextualised item type should be adopted for vocabulary size tests so that the test takers can respond to the required number of words within a reasonable period of time. Moreover, the definitions are expressed as synonyms or short phrases, to minimize the demands of the test task for the learners (Read, 1997).

The Eurocentres Vocabulary Size Test is a typical checklist test which presents learners with a series of words and simply requires them to indicate whether they know each one or not. As Read (1997) points out, since the format depends purely on self-reporting, there is an obvious problem with differing interpretations of what “knowing a word” means, as well as a lack of any means to check whether the learners are overestimating their vocabulary knowledge.

2.3 Definition and Assessment of Syntactic Knowledge
Following Urquhart and Weir, grammar is used in the traditional sense, to refer to syntax and morphology. For Sampson (1975: 38), syntax is defined as “how words are
put together to form sentences”. In Richards et al. (1992: 161), it is “the way in which linguistic units such as words and phrases are combined to produce sentences in the language”. To these researchers, syntactic knowledge primarily concerns the well-formedness (or ill-formedness) of a sentence or subparts of a sentence such as a clause or a phrase. However, Horrocks (1987), Crystal (1997) and Purpura (2004) attached more importance to meaning. Even in the brief definition of syntactic knowledge, there appears to be some significant divergences. Urquhart and Weir (1998: 259) remarked, “We still are faced with the problem of dividing on the scope of syntax, i.e. what comes under ‘syntax’ and is therefore to be included in our research, and what is outside, …where to place the boundary between syntax and semantics…” Their advice for the research of the contribution of knowledge of syntax to reading is to take a “formal”, “structuralism” model of syntactic knowledge, with as little recourse to “meaning” or “communicative value” as possible; that is, the syntactic knowledge only concerns the permissible (grammatical) sequences of words or formatives in the sentences of whichever language is under study.

In the research of the contribution of syntactic knowledge to reading comprehension, Alderson (1993) made clear that syntactic knowledge and reading comprehension tests should be as separate from each other as possible. However, this is not easy to achieve, because most tests of syntactic knowledge involve the processing of written language. Since the definition of the syntactic knowledge adopted in the research of syntax’s contribution to reading is suggested to be sentence-bound, one may achieve some measure of difference between the testing of syntactic knowledge and that of reading comprehension by basing a syntactic knowledge test on decontextualized sentences or phrases. This means not using continuous text, as in the cloze procedure, or the gap-filling of continuous text.

3. Method

3.1 Participants
The participants were 68 non-English major sophomores in Anhui Medical University. Their scores on CET Band 4 ranged from 324 to 561, with a mean score of 464.3.

3.2 Instruments

3.2.1 Reading Comprehension Test
The reading comprehension test consisted of four passages of different styles and registers. All the passages were chosen from the CET Band 4 reading test of the years 1999, 2000, 2002 and 2003. The items following each passage required the synthesis of information across several sentences, and a correct answer could not be identified by simply locating or understanding a particular word or phrase. The response format was multiple-choice. Following is a sample test item:

Which of the following is the author’s view on the historical development of toys?
A) The craftsmanship in toy-making has remained essentially unchanged.
3.2.2 Vocabulary Breadth Test

The vocabulary breadth test was devised on the basis of the Vocabulary Levels Test (Nation, 1990; Schimitt et al., 2001), which used the format of word-definition matching. In this study, the participants shared the same first language, so the English definitions of the tested words were converted into their Chinese equivalents.

Thirty items were randomly selected from the vocabulary list of CET Band 4. The test focused on three types of content words, nouns, verbs and adjectives. This was in line with the practice that a vocabulary test usually focuses knowledge of content words (Read, 2000).

Altogether ten groups of words were randomly selected. Each group had six words, three target answers, the other three distracters. An example will help a lot.

1  butterfly
2  contrast  (   ) 同情
3  evidence  (   ) 治疗
4  frog     (   ) 蝴蝶
5  sympathy
6  treatment

3.2.3 Syntactic Knowledge Test

The syntactic knowledge test used a multiple-choice format. Each item in the test presented a sentence with one part replaced by a blank, followed by four response options of similar semantic content; only one of the four satisfied the syntactic constraints imposed by the structure of the part of the sentence that was given.

The test only measured the knowledge of sentence structures tested in CET Band 4, so that the items corresponded to the test takers’ level of L2 proficiency. In accordance with the requirement of CET Band 4 and the statistical data collected by Wang (2001), 11 syntactic points were tested in CET Band 4, that is, agreement relationship, nominal clause, attributive clause, adverbial clause, it-sentence, there be-sentence, the use of “as”, inversion, emphasis, negation and substitution.

Since Urquhart and Weir (1998) and Alderson (1993) pointed out that in testing syntax, the instruments should reflect as closely as possible the construct alone, it appears that to achieve independence from a reading test, a syntax test should keep semantic processing and contextualization to minimum so as to limit the focus to syntactic knowledge. To focus more on an awareness of acceptable sentence constructions and less on sentence semantics, a number of stages were completed to eliminate semantically based items from the original syntax test which consisted of 90 items selected from the
The Relative Significance of Vocabulary Breadth and Syntactic Knowledge in the Prediction of...

Vocabulary and Structure of CET Band 4 from years 1989 to 2003. Three college English teachers in China were asked to screen out any items where the focus was not clearly on syntactic knowledge. Specifically, they were asked to answer:

Which of the three areas does each item seem to be testing?

1) The test takers’ “knowledge of meanings of certain words and phrases” (Lexical-Semantic Knowledge).
2) The test takers’ “understanding of the meaning of the overall sentence” (Sentence Reading Comprehension).
3) The test takers’ “knowledge of sentence structures and that of acceptable sequences and forms of words in terms of syntax” (Syntactic Knowledge).

After careful scrutiny of the teachers’ feedback, 30 items were seen as meeting the criteria and were thus obtained (one for substitution, two for negation and three for each of the other nine syntactic points). Following is a sample test item:

It wasn’t such a good dinner _______ she had promised us.
A) that  B) which  C) as  D) what

3.3 Procedures
The researcher did the following:
1) Prepare 68 test paper booklets, each of which contains the three tests developed in the present study;
2) Before the test, read aloud to the participants all requirements and rules of the test;
3) Hand out the test paper booklets to the participants. After 75 minutes, collect all the 68 paper booklets.

3.4 Scoring
For the reading comprehension test, there were four reading texts, each followed by five test items. Two points were assigned to each item that was correctly chosen, and the full mark for the test was 40 points.

For the vocabulary breadth test and the syntactic knowledge test, one point was assigned to each item that was correctly chosen, and the full mark for each test was 30 points.
Incorrect answers were given 0 in all the three tests.

3.5 Data Analysis
To examine the relative significance of vocabulary breadth and syntactic knowledge in the prediction of reading comprehension test performance, a multiple regression procedure was applied (with the significance level of 0.05) in which the dependent variable was reading comprehension test score and the independent variables were vocabulary breadth and syntactic knowledge test scores.
4. Results and Discussion

4.1 The Relative Significance of Vocabulary Breadth and Syntactic Knowledge in the Prediction of Reading Comprehension Performance

Table 1. Model summary of the multiple regression

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.551*</td>
<td>.304</td>
<td>.282</td>
<td>5.82681</td>
</tr>
</tbody>
</table>

Table 2. ANOVA analysis of the multiple regression

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>962.260</td>
<td>2</td>
<td>481.130</td>
<td>14.171</td>
<td>.000*</td>
</tr>
<tr>
<td>Residual</td>
<td>2206.858</td>
<td>65</td>
<td>33.952</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Coefficients of the multiple regression

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>3.301</td>
<td>4.327</td>
<td>.763</td>
<td>.448</td>
</tr>
<tr>
<td>syntactic knowledge</td>
<td>.537</td>
<td>.168</td>
<td>.368</td>
<td>.002</td>
</tr>
<tr>
<td>vocabulary breadth</td>
<td>.444</td>
<td>.184</td>
<td>.279</td>
<td>.018</td>
</tr>
</tbody>
</table>

Both independent variables, vocabulary breadth and syntactic knowledge, entered the regression model. F(2, 65) = 14.171, p = 0.000 (Table 2), and the t statistics of the two independent variables shown in Table 3 (3.195 and 2.417) were above +2, and their significance levels (0.002 and 0.018) were smaller than the P level of the present study (0.05). The analysis showed a positive linear relationship between the two predictors and the independent variable. That is to say, both vocabulary breadth and syntactic knowledge can predict the variation in reading comprehension. To be specific, 55.1% of variance in EFL learners’ reading comprehension (Table 1) can be predicted by their L2 vocabulary breadth and L2 syntactic knowledge.

Moreover, as for the relative significance of the two predictors, Table 3 reveals that syntactic knowledge exceeds vocabulary breadth in terms of standardized regression weights (0.368>0.279) and thus has stronger predicting power than vocabulary breadth.

L2 vocabulary and syntactic knowledge are considered to be the most important factors that affect L2 reading comprehension (Bossers, 1992; Brisbois, 1995; Nation & Coady, 1988; Ruddell, 1994; Tailleffer, 1996). Thus, it is reasonable to find that L2 learners’ vocabulary breadth and syntactic knowledge can predict their performance on L2 reading comprehension.

Since reading texts are composed of a number of words, knowledge of word meanings is related to the ability to comprehend a text. Obviously, if there are too many
unknown words in a text, it is easy to be lost in the text. A certain amount of vocabulary is considered to be the prerequisite for comprehending a reading text. For instance, Nation (1990) has found that 3,000 word family level or knowing 95% of the words appearing in a text is the minimum for the comprehension of a not very difficult text. In addition, it is asserted by Beck, Perfetti and Mckeown (1982) that the increase of vocabulary breadth can result in the improvement of reading comprehension. Vocabulary breadth affects reading comprehension in terms of the richness of a reader’s semantic representations. The L2 learners who possess a rich and interconnected lexical knowledge will comprehend a reading text better than those who do not. Spilich et al. (1979) have found that prior knowledge about the topic of a text can facilitate reading comprehension. Thus, if word meanings are poorly represented in semantic memory, less information will be accessed and then fewer relations between concepts will be made than if a rich semantic representation for word meanings exists.

However, limited vocabulary knowledge does not always impair reading comprehension. On the contrary, vocabulary knowledge per se seems to be insufficient to ensure adequate comprehension of larger units of text (Pany, Jenkins & Schreck, 1982). Although various skills can help the learners read effectively, Goodluck (2000) argues that it is the analysis of syntactic structures that helps them achieve the most efficient reading. Nuttall (2002) agrees with Goodluck (2000) by asserting that one of the main difficulties in reading comprehension is caused by various complicated syntactic structures and therefore, even if the learners are familiar with the vocabulary of the reading text, they will still be hindered by a lack of syntactic knowledge from comprehending the whole text. Many other researchers share the same view that readers may experience text comprehension difficulties even when vocabulary knowledge is controlled for (Ehrlich & Remond, 1997; Oakhill, Cain & Yuill, 1998; Stothard & Hulme, 1992). Grellet (1987) also emphasizes that learners should be able to analyze various syntactic structures and not be restrained by single words so as to read more efficiently. That is because even after vocabulary ability has been taken into consideration, syntactic knowledge is related to both components of reading ability, that is, reading comprehension and decoding (Willows & Ryan, 1986). It is suggested by Tunmer (1989) that there might be a reciprocal relation between reading ability and syntactic ability in that syntactic knowledge predicts both word decoding and listening comprehension, and these two skills in turn predict reading comprehension. Therefore, it is justifiable that, compared with vocabulary, syntactic knowledge appears to exert more effect on reading and is more significant in predicting reading comprehension test performance.

4.2 The Relationship Between EFL Learners’ L2 Proficiency and the Reading Comprehension’s Predictability by Vocabulary Breadth and Syntactic Knowledge

To examine the relationship between the participants’ English proficiency and the reading comprehension’s predictability by vocabulary breadth and syntactic knowledge, the participants were divided into two groups based on their CET Band 4 score. The top 22 candidates whose CET-4 scores ranged from 499 to 561 (with a mean of 524.1) were
selected as the high-proficiency learners, and the bottom 22 whose CET-4 scores ranged from 349 to 427 (with a mean of 402.5) were selected as the low-proficiency learners. To investigate whether the reading comprehension’s predictability by vocabulary breadth and syntactic knowledge changes with L2 proficiency, multiple regression procedures were applied for the two groups respectively.

**Table 4.** ANOVA analysis of the multiple regression for high-proficiency candidates

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>474.619</td>
<td>2</td>
<td>237.309</td>
<td>21.404</td>
<td>.000*</td>
</tr>
<tr>
<td>Residual</td>
<td>210.654</td>
<td>19</td>
<td>11.087</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the high-proficiency group, $F(2, 19) = 21.404, p = 0.000$ (Table 4), and the $t$ statistics of the two independent variables in Table 5 (2.334 and 3.690) were above +2 and their significance levels (0.031 and 0.002), were smaller than 0.05. Therefore, it showed that, as for the high-proficiency learners, their reading comprehension test performance can be effectively predicted by their vocabulary breadth and syntactic knowledge. However, as for the relative significance of the two predictors, Table 5 shows that syntactic knowledge is a better predictor of reading comprehension test performance than vocabulary breadth in terms of standardized regression weights (0.571>0.361).

**Table 5.** Coefficients of the multiple regression for high-proficiency candidates

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>-4.730</td>
<td>7.218</td>
<td>-.655</td>
<td>.520</td>
</tr>
<tr>
<td>vocabulary breadth</td>
<td>.728</td>
<td>.312</td>
<td>.361</td>
<td>2.334</td>
</tr>
<tr>
<td>syntactic knowledge</td>
<td>.668</td>
<td>.181</td>
<td>.571</td>
<td>3.690</td>
</tr>
</tbody>
</table>

**Table 6.** ANOVA analysis of the multiple regression for low-proficiency candidates

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>621.022</td>
<td>2</td>
<td>310.511</td>
<td>15.361</td>
<td>.000*</td>
</tr>
<tr>
<td>Residual</td>
<td>384.069</td>
<td>19</td>
<td>20.214</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 7.** Coefficients of the multiple regression for low-proficiency candidates

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>(Constant)</td>
<td>-4.126</td>
<td>4.178</td>
<td>-.987</td>
<td>.336</td>
</tr>
<tr>
<td>vocabulary breadth</td>
<td>.447</td>
<td>.210</td>
<td>.400</td>
<td>2.126</td>
</tr>
<tr>
<td>syntactic knowledge</td>
<td>.980</td>
<td>.398</td>
<td>.463</td>
<td>2.463</td>
</tr>
</tbody>
</table>
For the high-proficiency group, $F(2, 19) = 15.361$, $p = 0.000$ (Table 6), and the t statistics of the two independent variables in Table 7 (2.126 and 2.463) were above +2 and their significance values (0.047 and 0.024) were also smaller than 0.05; that is to say, vocabulary and syntactic knowledge take a significant part in predicting individual differences in the performance in a reading comprehension test. As for the relative significance of the two predictors, Table 7 has shown that syntactic knowledge exceeds vocabulary breadth in terms of standardized coefficients (0.463>0.400).

To sum up, L2 proficiency does not affect the predictability by the two independent variables. Not only the single group analyses with the full samples but also the subgroup analyses that considered learners’ different L2 proficiency showed that both vocabulary breadth and syntactic knowledge contribute to accounting for the variation in reading comprehension but syntactic knowledge is a stronger predictor.

No matter what the EFL learner’s L2 proficiency is, his or her reading comprehension can be more effectively predicted by L2 syntactic knowledge than by L2 vocabulary breadth. This is because syntactic structures closely interrelate with the information conveyed by the language (Rutherford, 1987). To be specific, the more complicated the information is, the more complex syntactic structure is usually used to convey it. Therefore, text reading comprehension, to some extent, relies on the analysis of syntactic structures. Oakhill et al. (2003) concluded that sufficient syntactic knowledge helps to improve reading comprehension in two perspectives. First, if readers are more sensitive to various syntactic structures, they are more likely to be able to infer, according to the context, the meaning of the new words appearing in the reading text, because “syntactic knowledge can aid word recognition if the readers can use the constraints of the sentence structure to supplement their decoding” (p. 444). Thus, learners with limited vocabulary can also efficiently comprehend a reading text if they have sufficient syntactic knowledge which can help them infer the meaning of the new words. Second, syntactic knowledge is correlated with reading comprehension, because understanding at the level of the sentence is obviously fundamental to understanding at higher levels. That is, the understanding of sentence structures is the basis of the comprehension of the whole text. More specifically, syntactic awareness can help readers detect and correct the mistakes in reading comprehension, and thus improve their proficiency of comprehension monitoring.

The results of the present study, though consistent with qualitative research in L2 reading (Bernhardt, 1999), are in conflict with most of the previous quantitative research investigating this issue (Barnett, 1986; Brisbois, 1995; Yamashita, 1999). It is necessary to revisit these studies to explore possible reasons for the divergence in findings. Barnett (1986) used a recall task to measure reading comprehension. However, it should be noticed that the requirement of memory in the recall task may hinder the test takers’ ability to demonstrate their comprehension of the reading passage (Chang, 2006), because readers may not be able to remember all they have understood. As for Brisbois (1995), though she reported that vocabulary correlated more strongly with reading than did syntax, some aspects of her method may have given some advantages to the vocabulary variable as a predictor of reading comprehension. For one thing, there is the issue of timing of her data collection. Her students took the grammar test two months before the collection of their
reading comprehension data, but they took the vocabulary test two weeks prior to the reading comprehension test. For the other, there is a lack of consistency: her grammar test was a general assessment, whereas the vocabulary test focused on some specific linguistic elements of the target reading passages. In Yamashita’s study (1999), which was based on Japanese university EFL learners, a gap-filling test was used to measure the reading comprehension. It can be noticed that when the part of her reading comprehension assessment based on multiple-choice questions was isolated as the criterion, the data did not demonstrate any substantive advantage of vocabulary over syntax. The data of these studies cannot be interpreted as clear evidence that vocabulary breadth accounts for significantly more variation in reading comprehension than syntactic knowledge does. Here the possible effects of the test method in assessing reading comprehension should be addressed. Since the present study was based on multiple-choice questions alone, the conservative interpretation of the findings is that, for the Chinese college EFL learners, syntactic knowledge plays a larger part than vocabulary breadth in explaining the learners’ individual differences in a multiple-choice based reading comprehension test.

5. Conclusion

In summary, on the basis of the componential model of reading, Nation’s construct of the vocabulary breadth test and structuralists’ view of syntactic knowledge, the present study investigated the relative significance of vocabulary breadth and syntactic knowledge in the prediction of reading comprehension performance. It was found that: 1) there is a positive linear correlation between the learners’ performance on the reading comprehension test and their vocabulary breadth and syntactic knowledge, with the multiple \( r \) being 0.551; 2) syntactic knowledge outperforms vocabulary breadth in predictive power and emerges as a stronger predictor of reading comprehension; and the relative significance of syntactic knowledge over vocabulary breadth in the prediction of reading comprehension does not change with EFL learners’ proficiency.

The findings of the study lead one to question the notion prevalent in L2 teaching and learning that to enhance reading ability, increasing vocabulary size is more effective than developing syntactic knowledge. They also shed light on the fact that Chinese college EFL students will benefit a lot from the training of syntactic knowledge. Influenced by the schema theory, which has been very popular since the last century, syntactic analysis in the reading class is challenged. Much emphasis has been laid on vocabulary breadth training and little attention has been paid to syntactic knowledge. Introduction and exercises on vocabulary are more often provided by both textbooks and teachers in reading class. But the fact is that no matter how many words students have acquired, if they fail to understand the syntactic structures of the sentences, they are still unable to comprehend the reading text. Therefore, it is suggested that syntactic knowledge training should be an indispensible part of a college English reading course. Instead of being vocabulary-oriented, the reading class should keep a balance between the training of vocabulary and that of syntactic knowledge.
With respect to pedagogy, an additional implication of the present study is related to the remediation of reading problems. Learners’ reading problems need to be properly diagnosed to determine which kinds of knowledge they are lacking. Students who fail a reading test may have problems because they do not have sufficient vocabulary breadth or syntactic knowledge, or both. Since the remediation procedures for poor reading and poor comprehension are very different in kind, one needs to know which knowledge is lacking before effective remediation procedures can be implemented. The different significance of vocabulary breadth and syntactic knowledge in predicting reading comprehension test’s variance can be employed as the reference to the different measures we take to detect reading problems and then to enhance learners’ reading ability respectively.

References


(Copy editing: Joshua Lee)