Knowledge of Formulaic Sequences as a Predictor of Language Proficiency

Vahid Rafieya 1,2*

1International College of Liberal Arts, Yamanashi Gakuin University, Kofu, Yamanashi, Japan
2Department of Comparative Study of Cultures, Tsuru University, Tsuru, Yamanashi, Japan

Corresponding Author: Vahid Rafieyan, E-mail: rafieyanv@ygu.ac.jp

ARTICLE INFO

Article history
Received: November 14, 2017
Accepted: January 22, 2018
Published: March 01, 2018
Volume: 7 Issue: 2
Advance access: February 2018

Conflicts of interest: None
Funding: None

ABSTRACT

Formulaic sequences are assumed to play a vital role in foreign language learners’ speech fluency and language proficiency as they constitute a major part of foreign language learners’ linguistic repertoire. In this respect, the current study examined the relationship between knowledge of formulaic sequences and language proficiency to scrutinize the significance of knowledge of target language formulaic sequences in determining target language proficiency. The participants of the study were 45 Japanese learners of English as foreign language at three different levels of language proficiency: low-intermediate (18 participants), intermediate (12 participants), and high-intermediate (15 participants) at the Intensive English Program of Yamanashi Gakuin University. The instrument used for data collection consisted of a 30-item oral-production discourse completion task to test language learners’ knowledge of formulaic sequences. The analysis of Spearman rank-order correlation coefficient (rho) revealed a strong positive relationship between language learners’ knowledge of target language formulaic sequences and their level of language proficiency. The pedagogical implications of the study suggested incorporation of target language formulaic sequences in every foreign language classroom instruction.

Key words: English as Foreign Language, Formulaic Sequences, Language Proficiency, Multiple-word Strings, Speech Fluency

INTRODUCTION

Formulaic sequences, a cover term for multiple-word strings that behave as single units such as collocations, idioms, proverbs, and lexical bundles (Boers et al., 2006; Alali & Schmitt, 2012; Grami & Alkazemi, 2016), constitute a major part of foreign language learners’ linguistic repertoire (Erman & Warren, 2000; Foster, 2001; Schmitt & Carter, 2004). The analysis of the rate of formulaic sequences in the spoken and written discourse of English by Erman and Warren (2000) revealed that 58.6 percent of the spoken English discourse and 52.3 percent of the written discourse was made up of formulaic sequences. The analysis of transcripts of unplanned English native speech by Foster (2001) also showed that 32.3 percent of unplanned English native speaker speech consists of formulaic language. The high rate of formulaic sequences in language indicates that they play a vital role in foreign language learners’ speech fluency and language proficiency (Boers et al., 2006; Schmitt, 2010; Stengers et al., 2011; Peters & Pauwels, 2015). Therefore, “it is impossible to perform at a level acceptable to native users, in writing or in speech, without controlling an appropriate range of multi-word units” (Cowie, 1992: 10). However, appropriate use of target language formulaic sequences has been found to be a particular challenge to foreign language learners, even at advanced levels (Nesselhauf, 2003; Durrant & Schmitt, 2009; Laufer & Waldman, 2011; Levitzky-Aviad & Laufer, 2013).

The widespread discussions over the significance of knowledge of target language formulaic sequences in determining target language proficiency on one hand and language learners’ underuse of these multiple-word strings on the other hand encouraged a number of researchers to investigate the relationship between knowledge of formulaic sequences and language proficiency. In one of these investigations, Bardovi-Harlig et al. (2008) explored the production of conventional expressions, a subgroup of formulaic sequences, by English as foreign language learners from different language-culture backgrounds at different levels of classes at the Intensive English Program of a university in the United States. The study consisted of 108 participants from four first languages of Arabic, Chinese, Japanese, and Korean. The data was collected through employment of a speech production task via a computer-delivered aural discourse completion task including six aural scenarios: two of thanking, two of apology, and two of refusal scenarios. The comparison of multiple first languages showed that learners of various first languages often share production strategies.

The comparison of different levels of language proficiency...
also showed that language learners increase their use of conventional expressions at higher levels requiring both linguistic and sociopragmatic competence. The findings of the study indicated that knowledge of formulaic sequences, regardless of language-cultural background, plays a significant role in determining the proficiency level of learners of English as a foreign language.

Qi and Ding (2011) were another group of researchers who conducted a study to investigate whether there is any improvement in formulaic sequence use in terms of frequency, accuracy, and variation in monologues of Chinese students majoring in English. In their study, they compared the spoken corpora between students of years one and four. The corpus was derived from the Longitudinal Spoken English Corpus of Chinese Learners which contained transcriptions of audio recordings of Chinese university English majors’ monologues on different topics. The data were analyzed qualitatively and quantitatively. Qualitative analysis involved the examination of the use of a particular formulaic sequence or a particular type of formulaic sequences. It also involved the examination of erroneous formulaic sequences together with the concordance lines in which they occurred in attempt to tease out the sources of errors. The quantitative analyses measured the frequency, accuracy, and variation of formulaic sequences when investigating the use of formulaic sequences in oral texts. The study revealed that there were no significant differences between the performance of students of year one and year four in formulaic sequence frequency and accuracy in their monologues, but there was significant improvement in formulaic sequence variation. The findings obtained by the study indicated that more proficient language learners possess a higher inventory of formulaic sequence knowledge.

In the same vein, Staples et al. (2013) conducted a study to determine whether the frequency, function, and fixedness of lexical bundles, a subgroup of formulaic sequences, used by English language learners in a controlled environment vary across proficiency levels and to see how language learners’ use of lexical bundles compares with that generally found in academic writing. The study used data from a corpus composed of written responses to items on the TOEFL iBT. The corpus contained two written texts from 480 participants. The responses were scored on a five-point scale in half point increments. The corpus was further subdivided into three proficiency levels (low, medium, and high) based on a range of Education Testing Service (ETS) scores. Biber, Conrad, and Cortes’ (2004) taxonomy was used to identify bundle functions. Following Biber (2009), the degree of fixedness for each of the four slots in the bundle was investigated in relation to the other three. The results indicated that lower level language learners used more bundles overall but also more bundles identical to those in the prompts. In contrast, the functional analysis revealed a similar use of stance and discourse organizing bundles across proficiency levels and very few referential bundles used by any of the groups. In addition, there were few differences in fixed versus variable slot bundles across proficiency levels. The findings of the study, unlike those obtained by Bardovi-Harlig et al. (2008) and Qi and Ding (2011), suggested a lack of correlation between knowledge of formulaic sequences and language proficiency.

In another attempt to examine the relationship between knowledge of formulaic sequences and language proficiency, Serrano et al. (2015) investigated whether intensive exposure is more beneficial for lower-proficiency language learners, or whether a certain command of the target language is necessary before language learners can benefit from intensive exposure to the language. Participants of the study included 124 Spanish/Catalan learners of English as foreign language enrolled in two program types: intensive and regular. Both programs offered 110 hours of English instruction distributed over four and a half weeks in the former and over 7 months in the latter program. The methodological approach, textbooks, and exams were the same for both programs, the only difference being time distribution. Three different proficiency levels of beginners, intermediate, and advanced were considered, as determined by their class level and on the basis of a range of independent proficiency measures in terms of complexity, accuracy and fluency. In order to examine language learners’ use of formulaic sequences, language learners’ target language performance was analyzed in an oral narrative based on a series of pictures. Language learners were given around 30 seconds to become familiar with the pictures and when they were ready they narrated the story. The results of the study suggested that concentrating time distribution of target language hours of instruction fosters the acquisition of formulaic sequences but only under certain conditions. Also, it was concluded that intensity is not equally beneficial for the acquisition of formulaic sequences at all proficiency levels. Advanced language learners did not seem to benefit from intensive instruction to the same extent as low-proficiency language learners. The findings of the study, similar to the findings of the study by Staples et al. (2013), showed that knowledge of formulaic sequences is not a predictor of language proficiency.

The review of literature, despite the widespread discussions raised over the significance of possessing a good command of formulaic sequences in developing proficiency in the target language and achieving native-like fluency by many scholars in the field of linguistics (e.g. Boers et al., 2006; Schmitt, 2010; Stengers et al., 2011; Peters & Pauwels, 2015), shows that only very few studies have investigated the use of formulaic sequences by language learners at different proficiency levels (Staples et al., 2013). In addition, these studies have revealed mixed findings. While some of the studies (e.g. Bardovi-Harlig et al., 2008; Qi & Ding, 2011) displayed a significant relationship between language learners’ knowledge of formulaic sequences and their language proficiency, some other studies (e.g. Staples et al., 2013; Serrano et al., 2015) found that knowledge of target language formulaic sequences does not lead language learners to higher target language proficiency. Thus, more studies need to be conducted to depict a clearer picture of the relationship between knowledge of formulaic sequences and language proficiency as it is unclear whether lower proficiency language learners use more or fewer formulaic
sequences than higher proficiency language learners (Boers et al., 2006; Forsberg, 2010). To this end, the current study seeks to investigate the relationship between knowledge of formulaic sequences and language proficiency. Therefore, the research question to be addressed in the current study is:

Is there any relationship between knowledge of formulaic sequences and language proficiency?

Accordingly, the null hypothesis is:

There is no relationship between knowledge of formulaic sequences and language proficiency.

METHODOLOGY

Participants

Participants of the study were 45 Japanese learners of English as foreign language at the Intensive English Program of International College of Liberal Arts, Yamanashi Gakuin University. Twenty-eight of the participants were males and the remaining 17 were females. Their ages ranged from 18 to 22, with a mean age of 19.2. The participants were at three different levels of language proficiency based on a TOEFL placement test: low-intermediate (18 participants), intermediate (12 participants), and high-intermediate (15 participants). None of the participants had the experience of living in an English speaking country over an extended period of time. However, they were taught by native English teachers and lived in the same dormitory with their international peers, enjoying an immersive English environment. Also, participants of all different levels of language proficiency were instructed based on the same teaching method and hours of instruction. Therefore, they were all equally exposed to English language and culture both inside and outside classroom environment.

Instrument

To test language learners’ knowledge of target language formulaic sequences, an oral-production discourse completion task developed by Bardovi-Harlig et al. (2015) was adopted. The test included two examples, two practice items, and 30 test items. The 30 items included 10 agreement, 10 disagreement, and 10 clarification (5 self-clarification and 5 other-clarification) scenarios. Each item started with a brief description of the topic and then for agreements and disagreements gave language learners a specific opinion. Language learners read the descriptions and their position. After language learners heard and read the setting and their position, they heard a classmate’s turn to which they responded. The narrator and the classmate alternated between a male and female voice so that language learners could easily distinguish the classmate’s turn from the narrator’s turn.

Data Analysis

To assess language learners’ performance on the discourse completion task, the responses to the items on the discourse completion task were transcribed by the researcher. The responses were then coded by two native English speakers. If a student produced the targeted speech act appropriately, the response received 1 point. However, if the targeted speech act was not produced appropriately, it earned no points. There were 10 points possible for expressions that occurred in the context of agreements, 10 points possible for expressions that occurred in the context of disagreements, and 5 points for each of self- and other-clarifications. As there were overall 30 items on the discourse completion task, each language learner could get a score ranging from 0 to 30. The percentage of items receiving 1 point for a targeted expression were calculated for each student for agreements, disagreements, and self- and other-clarifications. Inter-rater reliability for speech act identification was 92 percent.

To assess the relationship between language learners’ knowledge of formulaic sequences and their language proficiency, Spearman rank-order correlation coefficient (rho), which is used to test for a rank order relationship between two quantitative variables when concerned that one or both variables is ordinal (rather than interval) and/or not normally distributed or when the sample size is small (Gravetter & Wallnau, 2013), was used. The value of Spearman rank-order correlation coefficient (rho) can range from -1.00 to +1.00. The positive and negative signs out the front indicate whether there is a positive correlation (as one variable increases, the other variable increases as well) or a negative correlation (as one variable increases, the other variable decreases). The size of the value, regardless of the sign, provides an indication of the strength of the relationship. Values of closer to +1.00 or -1.00 are indicative of higher correlation between the two variables, whereas values of closer to 0.00 are indicative of a lower correlation (Pallant, 2013). Cohen (1988) suggests a set of guidelines to interpret the values between 0.00 and 1.00. The guidelines, which have been presented in Table 1, apply whether or not there is a positive or negative sign out the front of the correlation value.

The squared correlation ($r^2$), called the coefficient of determination, was then used to measure the proportion of variability in language proficiency that can be determined from its relationship with knowledge of formulaic sequences. Squared correlation would give a value ranging from 0.00 to 1.00. Cohen (1988) has also suggested a set of guidelines to interpret the values of squared correlation. The criterion for interpreting the value of squared correlation ($r^2$), as proposed by Cohen (1988), has been presented in Table 2.

FINDINGS AND DISCUSSION

Findings

Table 3 presents the descriptive analysis of knowledge of target language formulaic sequences for language learners both at each proficiency level and in general. As the data presented in the table shows, knowledge of target language formulaic sequences increases with proficiency level: language learners at low-intermediate level displayed the lowest knowledge (mean score: 9.61), language learners at intermediate level outperformed the low-intermediate ones but their performance was not as good as the high-intermediate counterparts (mean score: 13.33), and language learners at
Knowledge of Formulaic Sequences as a Predictor of Language Proficiency

high-intermediate level had the highest performance (mean score: 18.33). However, the mean performance of language learners at all proficiency levels in general was not remarkable (mean score: 13.51). This indicates that despite being at a moderately high level of proficiency, language learners do not possess a high command of target language formulaic sequences in general.

Table 4 presents the results of Spearman rank-order correlation coefficient (rho) analysis for language learners’ knowledge of formulaic sequences and their level of language proficiency. The first thing to consider in correlation analysis is the direction of the relationship between the variables (knowledge of formulaic sequences and language proficiency). The data shows that there is a positive relationship between the two variables, that is, knowledge of formulaic sequences increases so too does the level of language proficiency. The second thing to consider in correlation analysis is the size of the value of the correlation coefficient. This value will indicate the strength of the relationship between the two variables (knowledge of formulaic sequences and language proficiency). The value of correlation coefficient obtained in the analysis of Spearman rank-order correlation coefficient (rho) is 0.73 which according to the guidelines proposed by Cohen (1988) to interpret the values of correlation coefficient suggests quite a strong relationship between knowledge of formulaic sequences and language proficiency.

To get an idea of how much variance the two variables (knowledge of formulaic sequences and language proficiency) share, the coefficient of determination was calculated. This can be obtained by squaring the correlation value. The value of coefficient of determination for the obtained correlation analysis is $r^2 = (0.73)^2 = 0.53$ which according to the guidelines proposed by Cohen (1988) to interpret the values of coefficient of determination suggests a very large correlation coefficient. To convert the value of coefficient of determination to ‘percentage of variance’, it was multiplied by 100, that is, $r^2 = (0.73)^2 \times 100 = 53$. This suggests that knowledge of target language formulaic sequences helps to explain 53 percent of the variance in language learners’ level of language proficiency. The performance of language learners at different language proficiency levels has been depicted in Figure 1.

**Discussion**

The study found that there is a strong positive relationship between language learners’ knowledge of target language formulaic sequences and their level of language proficiency. Language learners who possessed a higher level of language proficiency demonstrated a higher level of knowledge of target language formulaic sequences than language learners who possessed a lower level of language proficiency. Therefore, the null hypothesis of the study which states that there is no relationship between knowledge of formulaic sequences and language proficiency is rejected.

These findings can be explained through the fact that normal discourse, both written and spoken, contains large percentage of formulaic sequences, making up between one

---

**Table 1. Strength of relationship**

<table>
<thead>
<tr>
<th>Correlation value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10-0.29</td>
<td>Small correlation</td>
</tr>
<tr>
<td>0.30 – 0.49</td>
<td>Medium correlation</td>
</tr>
<tr>
<td>0.50 – 1.00</td>
<td>Large correlation</td>
</tr>
</tbody>
</table>

**Table 2. Percentage of variance explained, $r^2$**

<table>
<thead>
<tr>
<th>$r^2$ value</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>Small correlation</td>
</tr>
<tr>
<td>0.09</td>
<td>Medium correlation</td>
</tr>
<tr>
<td>0.25</td>
<td>Large correlation</td>
</tr>
</tbody>
</table>

**Table 3. Knowledge of formulaic sequences at different proficiency levels**

<table>
<thead>
<tr>
<th>Levels</th>
<th>Number of participants</th>
<th>Mean score</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-intermediate</td>
<td>18</td>
<td>9.61</td>
<td>3.696</td>
</tr>
<tr>
<td>Intermediate</td>
<td>12</td>
<td>13.33</td>
<td>3.846</td>
</tr>
<tr>
<td>High-intermediate</td>
<td>15</td>
<td>18.33</td>
<td>3.177</td>
</tr>
<tr>
<td>All levels in general</td>
<td>45</td>
<td>13.51</td>
<td>5.133</td>
</tr>
</tbody>
</table>

**Table 4. Relationship between language proficiency and knowledge of formulaic sequences**

<table>
<thead>
<tr>
<th>Knowledge of formulaic sequences</th>
<th>Proficiency level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman’s rho</td>
<td></td>
</tr>
<tr>
<td>Knowledge of formulaic sequences</td>
<td>1.000</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>-</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>45</td>
</tr>
<tr>
<td>Proficiency Level</td>
<td>0.729**</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.000</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>45</td>
</tr>
</tbody>
</table>

**Correlation is significant at the 0.01 level (2-tailed)**

**Figure 1. Knowledge of Formulaic Sequences at Different Proficiency Level**
third and one half of discourse (Conklin & Schmitt, 2012). More exposure to target language input, through reading and listening materials, leads language learners to encounter more target language formulaic sequences. Research (Raichle, 1998; McCrone, 1999) has shown that once the brain is familiar with a linguistic task, it is able to bypass the processing route that was used to learn it (Wray & Perkins, 2000). Consequently, the brain would make use of a relatively abundant resources by storing frequently occurring formulaic sequences which could then be easily retrieved and used without the need to compose them online through word selection and grammatical sequencing (Pawley & Syder, 1983). This helps language users who possess a higher knowledge of target language formulaic sequences to be more proficient.

Although language learners who had a higher knowledge of target language formulaic sequences were more proficient users of target language, they were not at native-like level and did not show optimal performance on the test. This can be explained though Wray and Perkins’s (2000) model. According to this model, unlike native speakers who are likely to have stored common word sequences holistically, that is, as single unanalyzed chunks which can be retrieved from memory as prefabricated units and as a result bypass the need to assemble the sequences word by word, adult language learners are less likely to have stored conventional word strings holistically. Accordingly, the most likely processing benefit that formulaic sequences confer on a language learner is that particular formulaic sequences may be encountered often enough that the associations between the component words become so strong that, on meeting or recalling part of the string, the language learner will recall the rest (Boers & Lindstromberg, 2012).

The findings obtained in the current study are consistent with the findings obtained in the study conducted by Bardovi-Harlig et al. (2008) who found that knowledge of formulaic sequences, regardless of language-cultural background, plays a significant role in determining the proficiency level of learners of English as a foreign language. The findings of the current study are also in line with the findings obtained in the study conducted by Qi and Ding (2011) who found that more proficient language learners possess a higher inventory of formulaic sequence knowledge. The findings obtained in the study, however, do not support the findings obtained in the studies conducted by Staples et al. (2013) and Serrano et al. (2015) who found a lack of correlation between knowledge of formulaic sequences and language proficiency.

CONCLUSION

The study investigated the relationship between knowledge of formulaic sequences and language proficiency, that is, whether a high level of knowledge of target language formulaic sequences leads to a high level of target language proficiency or not. The findings of the study revealed that there is a strong positive relationship between knowledge of formulaic sequences and language proficiency. Language learners at higher levels of target language proficiency demonstrated a better command of target language formulaic sequences than language learners at lower levels of target language proficiency. Therefore, target language formulaic sequences are advised to be an indispensable part of foreign language instruction (Rafieyan, 2015; Rafieyan, 2016a; Rafieyan, 2016b).

REFERENCES


Knowledge of Formulaic Sequences as a Predictor of Language Proficiency


