

# International Journal of English Language & Translation Studies

ISSN: 2308-5460



## A Pilot Study on EFL Learners' Spoken Fluency and the Formulaic Language Use

[PP: 88-98]

**Junlei, Xuan**

Xinyang Normal University, **China**  
Chonbuk National University, **Korea**

**Huifang, Yang**

(Corresponding Author)

Xinyang Normal University  
**China**

### ABSTRACT

This pilot study examines the relationship between EFL learners' spoken fluency and the use of two-word formulaic sequences and three-word lexical bundles in their English speaking performance. 24 third-year English majors from a central China university are the participants and their speech samples are collected based on an English-speaking task. The temporal indices of spoken fluency which consist of SR (speech rate), AR (articulation rate), MLR (mean length of run) and PTR (phonation time ratio) are extracted as the dependent variables. And the speech samples are transcribed and two linguistic variables of formulaic language use, F2R (two-word formulaic sequences/run ratio) and B3R (three-word lexical bundles/run ratio) are also extracted as the independent variables. A canonical correlation analysis (CCA) is thereafter conducted to investigate the relationship between the dependent variables and the independent variables. Results show that there is a significant relationship between the learners' spoken fluency and their use of formulaic language.

**Keywords:** *Formulaic Sequences, Lexical Bundles, EFL, Chinese Learners, Spoken English Fluency, Canonical Correlation Analysis*

| ARTICLE | The paper received on | Reviewed on | Accepted after revisions on |
|---------|-----------------------|-------------|-----------------------------|
| INFO    | 29/10/2019            | 28/11/2019  | 20/01/2020                  |

#### Suggested citation:

Cite this article as: Junlei, X. & Huifang, Y. (2019). A Pilot Study on EFL Learners' Spoken Fluency and the Formulaic Language Use. *International Journal of English Language & Translation Studies*. 7(4). 88-98.

## 1. Introduction

### 1.1 English Formulaic Language and Its Measure

What is English formulaic language? It seems that it has got various names along with a lot of definitions. According to Schmitt (2000), numerous terms that have been coined to refer to the multi-word unit, the most common used terms are lexical chunks and lexical phrases. Another well-known researcher, Wray (2002) also points out that there are over 40 terms such as routine formula, formulaic language, recurring utterances, multiword lexical phenomena, lexicalized sentence stems, fixed expressions etc. Regardless of the various terms, just as Weinert (1995) argued that even a variety of labels have been used to describe formulaic language, but it seems that researchers have very much the same phenomenon in mind.

According to Alarbai(2016), idioms, collocations, phrasal verbs, fixed expressions and lexical bundles are all considered as formulaic sequences. However, for EFL learners, such a definition

might sound a little vague, because as non-native English speakers, they may still have difficulty in recognizing and identifying the formulaic expressions in various discourses. Therefore, based on previous studies, particularly with reference to Wray (2008) and Le-Thi et.al (2017), we'd like to adopt this following definition in this research: *English formulaic language refers to any English formulaic expressions from two-word expressions (you know, I see, what's up...) to multi-word expressions (what's going on, be all ears, the thing is...), which are already institutionalized and frequently used in the English community.* Such formulaic expressions are generally considered as the basic building blocks of English discourse. According to this definition, English formulaic language consists of various formulaic expressions and formulaic sequences, and it either functions with its linguistics features or its pragmatic features on a daily basis.

The measure of formulaic language was adopted from Wood (2010), Huang (2012) and Quan (2016). Formula/Run Ratio

(FRR) was generally used to measure the EFL learners' use of formulaic language. According to Wood (2010), FRR is a quantitative measure of how the use of formulas contributed to longer runs. And FRR has been utilized as an indicator of the average number of accurately produced formulaic expressions per run. Based on this measure, Formula/run ratio (FRR) can be calculated with the total number of formulaic expressions divided by the total number of runs.

### 1.2 Spoken Fluency and Its Measures

Fluency in second language literature is often distinguished from accuracy, researchers have identified oral fluency as native-like rapidity, such as flow, continuity, automaticity, and smoothness of speech (Huang, 2012). In EFL learning, fluency is not a new term. But how to define fluency, that is a question. Linguists have given definitions from different perspectives, some focus on the specifics of fluency, while others tend to examine fluency as holistic impressions. Previous studies focus on defining fluency are exemplified as Craig Lambert and Judit Kormos (2014) and Thomson (2015), and studies emphasize on measuring fluency are like Lennon (1990) and Towell et al. (1996).

Segalowitz (2010) presented that fluency in second language acquisition could be categorized into three types, cognitive fluency, perceived fluency and utterance fluency. Since perceived fluency and cognitive fluency seem to be either subjective or hardly perceived, therefore, utterance fluency is generally considered to be addressed in second language acquisition because it can be measured with the objective acoustic features of an utterance. According to Wood (2006), research on fluency mainly focus on measurable temporal variables in speech such as speech rate, pause, the length of fluent runs of speech between pauses, which provided reliable measures for this study that could help to determine speech fluency.

In such studies on fluency, temporal variables such as SR (speech rate) and MLR (mean length of run) have been used because of their significant relationship with standardized proficiency tests (Quan, 2016). According to Wood(2006) that previous studies on fluency concentrated mainly on measurable temporal variables in speech such as SR (speech rate), AR (articulation rate), MLR(mean length of run) and PTR(phonation time ratio), which provided reliable measure to determine fluency in

speech production. In this study, the spoken fluency measures of the temporal indices were adopted from Wood (2010), De Jong and Perfetti (2011), Huang (2012) and Quan (2016):

1. Speech Rate (SR): Total number of syllables uttered in response time divided by the total response time, including pauses
2. Articulation Rate (AR): Total number of syllables divided by the phonation time, or the actual speaking time excluding pauses
3. Phonation Time Ratio (PTR): Total amount of speaking time divided by the total response time
4. Mean Length of Run (MLR): Total number of syllables divided by the total number of runs. Run boundaries were determined by filled pauses and unfilled pauses of 0.3seconds or greater.

### 1.3 Aim of this Study

Compared with reliable measures from the previous studies on fluency, it seems there is no consensus on the identification and categorization of English formulaic language based on the previous studies on formulaic language, which might be one of the reasons that there haven't been many studies on the relationship between English formulaic language and spoken fluency. Nevertheless, there have been a few studies which indicate the associations between spoken fluency and the use of formulaic language. One example is Thomson (2017, p.26) that argues by saying that "previous research has shown a link between the use of multiword expressions and spoken fluency". Another example can be seen in Wood (2015) that also presents that formulaic language maybe a key element of second language speech fluency. However, according to McGuire& Larson-Hall (2017), the amount of empirical research providing evidence that use of formulaic sequences improves second language oral fluency is still small. Therefore, this study is carried out to address this issue by further investigating the relationship between EFL learners' English spoken fluency and their use of English formulaic language in terms of two-word formulaic sequences and three-word lexical bundles.

### 1.4 The Research Questions

1. What are the distributions of English formulaic language in terms of two-word lexical phrases and three-word lexical bundles in EFL learners' English-speaking performance?



2. Is there a noteworthy relationship between the EFL learner's spoken English fluency and their use of two-word lexical phrases and three-word lexical bundles?

## **2. Methodology**

### **2.1 Participants**

24 third-year English majors were selected as the participants in a normal university of Henan province, central China for the pilot study, at the early time of the second semester of the 2018-2019 academic year. Since most of the English majors are female students in this university, for the 24 participants, there are only 2 male students and 22 female students. Such EFL learners are pre-intermediate English learners. Because only about half of them passed the TEM4 nine months ago (TEM4- a nationwide English proficiency test which is designed and implemented for English Majors after their two years' study).

### **2.2 Speaking Task**

Based on Underhill (1987, p.66) and Wood (2010, pp.101-102), pictures were chosen online for an English-speaking task, in terms of the content of each picture, it is easy to describe even though they are different from each other. The pictures were edited and put in three PowerPoint files. Each participant was asked to choose one of the PPT files and select one from eleven pictures for their 60-90 seconds English-speaking task, each of the participants had 30- 60 seconds to prepare for the English-speaking task.

### **2.3 Context**

Since the speaking task was carried out by one of the English teachers of the department, so the context was a multimedia classroom for an interpreting class. All the speech samples were recorded in class by the computers of the classroom and transferred to the teacher's computer later.

### **2.4 Transcription**

Speech samples were collected, since one of the twenty-four recordings got damaged, we finally had 23 speech samples. And the 23 speech samples were transcribed and checked by 3 teachers from the School of Foreign Languages of the university, since the main purpose of the transcription was for identifying and extracting the formulaic language categories, therefore the grammatical errors were originally kept. And we had 23 transcripts ready for further analyzing the participants' specific use of English formulaic language.

### **2.5 Tools and Software**

To process the speech samples and extract temporal features of participants'

English spoken fluency, *Format Factory* - a software for converting *various* audio formats to WAV format was used. Because only the audios with WAV format fits PRAAT. With the script provided by De Jong (2009), the speech samples with WAV format speech samples was processed by PRAAT automatically to extract the temporal indices of fluency measures. Excel was also used to save and organize data, and SPSS.23 was utilized to run CCA (canonical correlation analysis) to examine the relationship between those two sets of variables.

## **3. Data Collection & Analysis**

### **3.1 Measuring Formulaic Language**

It has always been difficult for the researchers in this field to identify and extract formulaic language due to a lack of standard method. Without categorizing formulaic language into groups, it might not be possible to do so. In the previous studies, researchers offered two options. One was to invite English native speakers as judges to identify English formulaic language such as Wood (2010) did, the other was to have access to corpus to identify English formulaic language on their frequency and MI score such as Huang (2012) and Russel (2017) did.

Compared to other studies identifying and categorizing of English formulaic language, we found two studies more beneficial for our research because they provided us with some instructional guidance. One of the studies is from Vilkaitė (2016), who has done much in this aspect by investigating the distributions of the formulaic language categories such as collocations, phrasal verbs, idiomatic phrases and lexical bundles in four English registers (academic prose, fiction, newspaper language and spoken conversation), the other is Garnier (2016), who not only categorized formulaic language into lexical phrases, lexical bundles, phrasal expressions, idioms, collocations, and phrasal verbs, but also provide empirical evidence about L2 learners' knowledge of phrasal verbs.

And based on Garnier (2016) and Vilkaitė (2016), this study explores the two types of English formulaic language with four categories, two types refer to the English formulaic language with two different lengths, namely, *two-word formulaic sequences or lexical phrases (hereby these two terms are interchangeably used in this study)* and *three-word lexical bundles*. Four categories are respectively

*collocations, phrasal verbs, idiomatic phrases and lexical bundles.* To avoid the overlapping of these different categories, we identify and extract the formulaic languages with an order of three-word lexical bundles, two-word formulaic sequences. Within the group of two-word formulaic sequences, the formulaic languages are identified and extracted with an order of two-word collocations, two-word phrasal verbs and two-word idiomatic phrases.

### 3.1.1 Identifying and extracting three-word lexical bundles

Based on the definition given by Biber. et. al. (1999), lexical bundles are identified as the combinations of words that in fact recur most commonly in a given register. To qualify as a lexical bundle, a lexical sequence must occur at least *ten time per million words* in a register. And these occurrences must be spread across *at least five different texts* in the register. To identify and extract the three-word lexical bundles, we turned to use *Compleat Lexical Tutor<sup>i</sup>*, with reference to BNC spoken(1 million words), the three-word bundles that qualify 10 hits / millions across five different texts in the transcripts were identified and extracted.

### 3.1.2 Identifying and extracting two-word formulaic sequences

Compared to the three-word lexical bundles, it was more difficult for us to identify and extract two-word formulaic sequences because we had to deal with three subcategories: two-word collocations, two-word phrasal verbs and two-word idiomatic phrases. Regarding these three subcategories, it is not easy to distinguish them from each other, particular to the two-word phrasal verbs and two-word idiomatic phrases, because there is no absolute distinction between them. There is no standard method to do that. However, we could find ways to differentiate them. On the basis of the previous studies about identifying and extracting formulaic language, in this research, we identified and extracted the three categories of English formulaic language within the two-word formulaic sequences with an order of two-word collocations, two-word phrasal verbs and two-word idiomatic phrases.

#### a) Identifying and Extracting Two-word Collocations

Collocation is not a new term. Wood (2015, p.4) mentioned that early research to collocations was initiated by Firth (1951, 1957) and there were generally two types of collocations. One is the habitual collocation,

in which words occur together frequently. The other is the idiosyncratic collocation, a co-occurrence of words that relatively happens and yet has a function. The overall approach to collocations was developed by researchers such as Halliday, Mitchell and Greenbaum, Sinclair and Kjellmer.

In terms of the definitions of collocations, researchers have attempted to illustrate the language phenomena in different ways. Biber et al. (1999) defined the collocations as the associations between lexical words so that the words co-occur more frequently than expected by chance. And Vilkaitė (2016) suggested that collocations are considered to be a very frequent and important part of English language. In linguistics, collocation is the way that some words occur regularly whenever another word is used<sup>ii</sup>. In this study, online Oxford collocation dictionary<sup>iii</sup> is used as the basis for identifying the extracted two-word collocations from the transcripts. Five types of collocations (Adv + adj, e.g. *very good*; Verb + adj e.g. *felt good*; Verb + adv, e.g. *live happily*; Verb + noun e.g. *fly kites*; Adj + noun; *old man*) are identified and correspondingly, the number of the two-word collocations used in the transcripts were extracted.

#### b) Identifying and Extracting two-word Phrasal Verbs and Idiomatic Phrases

Identifying and extracting two-word phrasal verbs were once a challenge for this study. In the end, based on the definitions of phrasal verbs of the previous studies, we found a way. One important study in this field is Garnier(2016), which held the view that phrasal verbs can be defined as “word combinations that consist of a verb and a morphologically invariable particle, such as *look up, make out, or go through*” Garnier (2016,p.30). Another significant study, we believe, is Vilkaitė (2016) as he argued that phrasal verbs can be defined as sequences of verbs and adverbial particles that carry a single meaning.

However, when it comes to identifying phrasal verbs, the former definition seems to be too general while the latter seems to be too specific. Therefore, we revised the definitions and adopted the revision as an operational definition for identifying two-word phrasal verbs in this research. And the revised definition is that phrasal verbs are two-word sequences of verbs and adverbial particles or verbs and prepositional particles that carry a single meaning. In addition to the lists of formulaic sequences provided by Garnier (2016) and Vilkaitė (2016), for



identifying phrasal verbs and idiomatic phrases, other frequent occurring idiomatic phrases and phrasal verbs from the previous studies (Shin et al,2008; Liu,2003; Liu,2011; Martinez & Schmitt,2012; Russel,2017) were also added. The two-word phrasal verbs and idiomatic phrases mentioned above were extracted and put together. Finally, we had a list of 618 two-word sequences after deleting the overlapping ones. Such a list of two-word sequences along with the operational definition of phrasal verbs, were used as the basis for identifying two-word phrasal verbs, when phrasal verbs were extracted and then we dealt with the extracting of two-word idiomatic phrases based on the list. Thus, the number of the two-word phrasal verbs and idiomatic phrases used in the transcripts were identified and extracted.

Table 1: Two indicators of formulaic language use of 5 speech samples

| Subject | Id  | F2R  | B3R  |
|---------|-----|------|------|
| 1       | 112 | 0.06 | 0.16 |
| 2       | 116 | 0.22 | 0.12 |
| 3       | 127 | 0.17 | 0.56 |
| 4       | 210 | 0.32 | 0.28 |
| 5       | 025 | 0.36 | 0.35 |

Note: F2R=two-word formulaic sequences/run B3R=three-word bundles/run

### 3.2 The Method of Measuring Spoken English Fluency of the EFL Learners

Since the spoken fluency measures by its temporal variables were adopted from Wood (2010), De Jong and Perfetti (2011), Huang (2012) and Quan (2016), the following temporal indices of spoken fluency such as SR, AR, PTR and MLR were extracted from the speech samples by using PRAAT and its script. Specifically, PRAAT can be utilized for extracting the temporal indices automatically with an application of the script package of PRAAT-Scripts-master. The script package can be downloaded from the website<sup>iv</sup>. To extract the temporal indices of a speech sample, you can open the sound file in PRAAT and then choose the script (Figure 1) from the script package and run the script named *praat-script-syllable-nuclei-v2file.praat*.

Praat Script Syllable Nuclei  
Copyright (C) 2008 Nivja de Jong and Ton Wempe  
source: <https://sites.google.com/site/speechrate/Home/praat>  
  
This program is free software: you can redistribute it and it under the terms of the GNU General Public License as the Free Software Foundation, either version 3 of the License (at your option) any later version.  
  
This program is distributed in the hope that it will be but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See GNU General Public License for more details.  
  
You should have received a copy of the GNU General Public License along with this program. If not, see <http://www.gnu.org>

Figure 1: The script of praat-script-syllable-nuclei-v2file.praat

Since the default value of the minimum pause duration is 0.4 seconds, we need to change 0.4 to 0.3(Figure 2), according to De Jong (2009), the minimum pause duration in PRAAT is usually defined as no less than 0.3 seconds. Also, you have to make sure that both the directory and the sound file name are right.

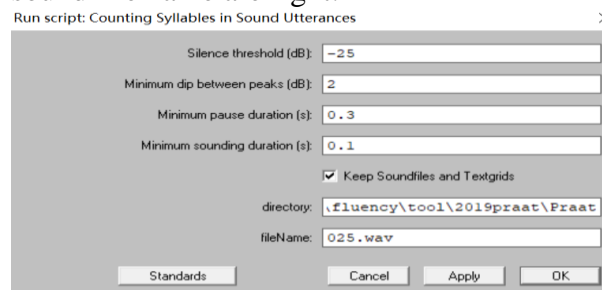


Figure 2: The running of the script and the change of minimum duration

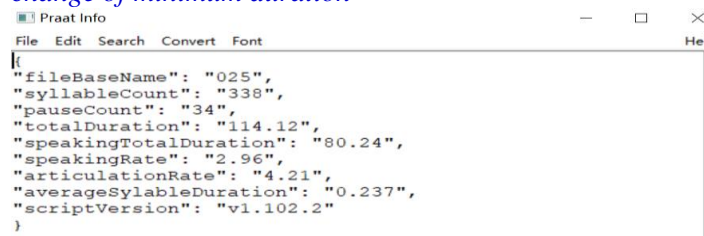


Figure 3: Praat information of the sound file

With the script provide by De Jong (2009), 23 speech samples were processed by PRAAT and the indices of the temporal features such as SR (speech rate), AR (articulation rate), MLR (mean length of run) and PTR (phonation time ration) of EFL spoken English fluency were saved in an Excel file with the order of the participants' student ID number.

Table 2: Four temporal indices of 5 speech samples

| Subject | Id  | SR   | AR   | MLR | PTR  |
|---------|-----|------|------|-----|------|
| 1       | 112 | 2.97 | 3.28 | 5   | 0.91 |
| 2       | 116 | 2.52 | 2.79 | 10  | 0.90 |
| 3       | 127 | 2.90 | 3.02 | 8   | 0.96 |
| 4       | 210 | 1.91 | 2.39 | 9   | 0.80 |
| 5       | 025 | 2.96 | 4.21 | 10  | 0.70 |

## 4. Data Analysis

### 4.1 The Distributions of the Formulaic Language in Speech Samples

Table 3, Table 4, Figure 4 and Figure 5 are all used to address the first research question: What are the distributions of the English formulaic language use in terms of two-word lexical phrases and three-word lexical bundles in the EFL learners' speaking performance?

It can be seen from Table 3 and Figure 4 that two types of English formulaic languages in the EFL learners' speech samples, namely, two-word formulaic sequences and three-word lexical bundles, have slightly different distributions, and three-word lexical bundles have been found

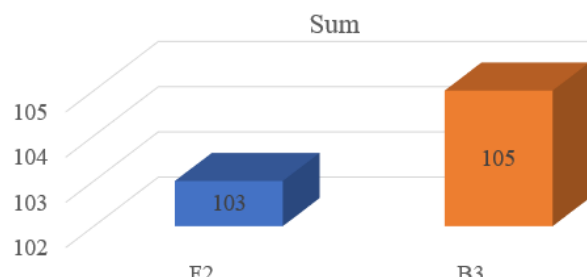
to be the more frequent-occurring formulaic language in the speech samples.

Both Table 4 and Figure 5 inform about the distributions of the four categories of formulaic language in the speech samples. And it should be noted that the four categories of formulaic language also have different distributions, with three-word lexical bundles being the most frequently used category, followed by two-word collocations and two-word idiomatic phrases. However, two-word phrasal verbs have been found to be the least frequently used formulaic language category.

*Table 3: Descriptive statistics of Linguistic variables of English formulaic language in speech samples*

| Variables          | N  | Minimum | Maximum | Sum    | Mean | Std. Deviation |
|--------------------|----|---------|---------|--------|------|----------------|
| F2                 | 23 | 1.00    | 10.00   | 103.00 | 4.48 | 2.23           |
| B3                 | 23 | 1.00    | 11.00   | 105.00 | 4.57 | 2.74           |
| Valid N (Likewise) | 23 |         |         |        |      |                |

Note: F2=two-word lexical phrases B3=three-word lexical bundles



*Figure 4: The distribution of two-word formulaic sequences and three-word lexical bundles in speech samples*

a) The distributions of two types of formulaic language in speech samples

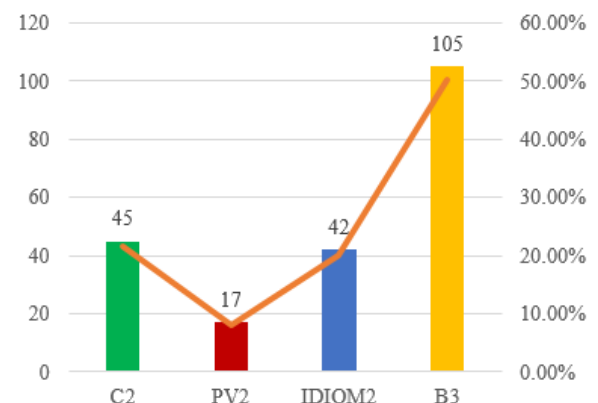
Both Table 3 and Figure 4 clearly show us that there is a slightly uneven distribution of two-word lexical phrases and three-word lexical bundles in the transcripts. Compared with two-word lexical phrases, three-word lexical bundles are a little more frequently used.

b) *The distributions of the four categories of formulaic language in speech samples*

As for the distributions of the four categories (three-word lexical bundles(B3), two-word collocations(C2), two-word phrasal verbs (PV2) and two-word idiomatic phrases (IDIOM2)) in the speech samples. It can be seen from Table 4 and Figure 2 that three-word lexical bundles have been found to be the most the most frequently used formulaic language, followed by two-word collocations and two-word idiomatic phrases, however, two-word phrasal verbs have been found to be the least frequently used formulaic language across the 23 speech samples.

*Table 4: Descriptive statistics about the distributions of the four formulaic language categories*

| FL Item            | N  | Minimum | Maximum | Sum    | Mean | Std. Deviation |
|--------------------|----|---------|---------|--------|------|----------------|
| C2                 | 23 | 0.00    | 6.00    | 45.00  | 1.96 | 1.46           |
| PV2                | 23 | 0.00    | 5.00    | 17.00  | 0.74 | 1.18           |
| IDIOM2             | 23 | 0.00    | 8.00    | 42.00  | 1.83 | 1.47           |
| B3                 | 23 | 1.00    | 11      | 105.00 | 4.57 | 2.74           |
| Valid N (Listwise) | 23 |         |         |        |      |                |



*Figure 5: The distributions of the four categories of formulaic language in speech samples*

Note: B3=three-word lexical bundles; C2=two-word collocations; PV2= two-word phrasal verbs; Idiom2=two-word idiomatic phrases

*4.2 Canonical Correlation Analysis (CCA) and its Interpretation*

Canonical correlation analysis (CCA) is a statistical technique which fits the study of relationships between multiple dependent and multiple independent variables (Mandal et al, 2017).

As a multivariate technique, CCA has several advantages. First, it limits the possibility of making Type One error. Second, a very important advantage of multivariate techniques such as CCA is that they may best capture the reality of psychological research. Third, this technique can be used in many instances, which makes it important and comprehensive as well (Sherry and Henson, 2005). Since we have two sets of variables, the dependent variables set (also the spoken fluency variables set that consists of four variables such as SR-speech rate, AR-articulation rate, MLR-mean length of run and PTR-phonation time ratio) , and the independent variables set (also the variables' set about the EFL learners' use of formulaic language that has two variables, namely, F2R, two-word formulaic sequences/runs ratio, and B3R, three-word lexical bundles/ runs ratio). Therefore, it is quite appropriate to employ the CCA (canonical correlation analysis) model in this research, the rationale can also be found in Sherry and Henson (2005), in which they hold the view that if researchers



have two variables sets in the study to examine their relationship, the use of CCA (canonical correlation analysis) is most appropriate. Furthermore, CCA (canonical correlation analysis) has been widely used in the fields of psychology, economics and speech recognition (Mandal et al. 2017).

Table 5: Bivariate correlations between the variables

|      |                     | Correlations |       |       |       |       |       |
|------|---------------------|--------------|-------|-------|-------|-------|-------|
|      |                     | ZSR          | ZAR   | ZMLU  | ZPTR  | ZF2R  | ZB3R  |
| ZSR  | Pearson Correlation | 1            | .931  | -.180 | .109  | -.400 | -.381 |
|      | Sig. (2-tailed)     |              | .000  | .412  | .621  | .058  | .073  |
| ZAR  | Pearson Correlation | .931         | 1     | -.262 | -.253 | -.363 | -.400 |
|      | Sig. (2-tailed)     | .000         |       | .226  | .244  | .088  | .058  |
| ZMLR | Pearson Correlation | -.180        | -.262 | 1     | .239  | .547  | .469  |
|      | Sig. (2-tailed)     | .412         | .226  |       | .273  | .007  | .024  |
| ZPTR | Pearson Correlation | .109         | -.253 | .239  | 1     | -.062 | .031  |
|      | Sig. (2-tailed)     | .621         | .244  | .273  |       | .777  | .888  |
| ZF2R | Pearson Correlation | -.400        | -.363 | .547  | -.062 | 1     | .331  |
|      | Sig. (2-tailed)     | .058         | .088  | .007  | .777  |       | .123  |
| ZB3R | Pearson Correlation | -.381        | -.400 | .469  | .031  | .331  | 1     |
|      | Sig. (2-tailed)     | .073         | .058  | .024  | .888  | .123  |       |

a. Listwise N=23

Tables 5 illustrates the bivariate correlations between the variables. ZSR and ZAR are highly correlated, therefore, for further applying of CCA, the variable ZSR was dropped, and ZAR was kept due to its nature of accuracy. Therefore, only 5 out of 6 original variables were used in the CCA for further analysis.

Table 6: Variable Sets for Canonical Correlation

| Dependent Variables' set | Independent Variables' set                  |
|--------------------------|---|
| AR=articulation rate     | F2R= two-word formulaic sequences/run ratio |
| MLR= mean length of run  | B3R=three-word lexical bundles/ run ratio   |
| PTR=phonation time ratio |   |

Table 6 illustrates the variables sets in the canonical correlation analysis (CCA). And it can be seen that the independent variables' set has two variables which are F2R (an indicator of the use of two-word formulaic sequences among the EFL learners) and B3R (an indicator of the use of three-word lexical bundles among the EFL learners). In terms of the dependent variables' set, now that we have three variables left of temporal indices of spoken fluency, which are respectively AR, MLR and PTR.

Table 7: Summary of Canonical Correlations Analysis

| Canonical Function | Canonical Correlation | Wilk's Lambda | Eigenvalue | DF | Sig. |
|--------------------|-----------------------|---------------|------------|----|------|
| 1                  | .74                   | .44           | 1.22       | 6  | .018 |
| 2                  | .11                   | .99           | .01        | 2  | .898 |

\*P<0.05

Tables 7 provides the summary of canonical correlation analysis. Results show that function 1 is the only statistically significant function (Rc=.74 with p<0.05). There is probably a relationship between the EFL learners' spoken fluency their use of

two-word lexical phrases and three-word lexical bundles. It should be noted that statistically significant results might be impacted heavily by sample size, therefore, it is important to interpret effect size indices to determine the CCA model's practical significance (Ho, 2014). The results show that the effect size is .56 (1-Wilk'S).

Table 8: Proportion of Variance Explained

| Canonical Variable | Set 1 by Self | Set 1 by Set 2 | Set 2 by Self | Set 2 by Set 1 |
|--------------------|---------------|----------------|---------------|----------------|
| 1                  | .368          | .203           | .663          | .365           |
| 2                  | .348          | .004           | .337          | .004           |

Table 8 illustrates the proportion of variance explained by the canonical variates, the results show that there is about 20% amount of variance of spoken fluency variables set can be predicted or explained by the linguistics variables' set of formulaic language use.

It can be concluded from the canonical correlation analysis so far, there is indeed a noteworthy canonical relationship between the EFL learners' spoken English fluency and their use of two-word formulaic sequences and three-word lexical bundles. This conclusion is based on a statistical significance of the relationship (Rc=0.74 with p<0.05), an effect size (0.56) and the amount of variance explained (20.3%) by canonical function 1.

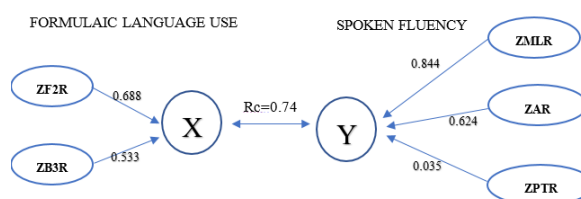


Figure 6: The top ranked variables in the first canonical function and the absolute value of their canonical loadings. A multivariate correlation of Rc=0.74 (\*p value<0.05)

Figure 6 illustrates the direct contribution of the variables to the canonical variates. The results show that the three variables of the fluency measures do contribute to the Canonical Variate Y, with ZMLR (mean length of run) being the top ranked variable which contributes most to the fluency composite, followed ZAR (articulation rate). But ZPTR (phonation time ratio) is found to contribute the least to the fluency composite, indicating the variable doesn't appear to be related to the canonical variate of spoken fluency.

As for the predictor variables, the two linguistic variables of the English formulaic languages, ZF2R (two-word formulaic

sequences/run ratio) and ZB3R (three-word lexical bundles/run ratio) have different contributions to the formulaic language composite X. ZF2R has been found to be the variable that contributes more to the Canonical Variate X. The results indicate that as far as the two categories of English formulaic language are concerned, the less frequently used two-word formulaic sequences (F2R), however, have been found to be a more important factor that contributes to the EFL learners' spoken fluency.

Therefore, both two forms of formulaic languages have made different contributions to the significant relationship extracted, compared with the more frequently used three-word lexical bundles, the less frequently used formulaic language, two-word formulaic sequences might matter more to facilitate EFL learners' spoken English fluency.

## 5. Discussion & Conclusion

To investigate the relationship between the use of English formulaic language and the spoken English fluency, 24 pre-intermediate Chinese EFL learners were selected, an English-speaking task was assigned to them and the recording of the speech samples were transcribed and checked. The audio data and the text data were collected. As for the spoken English fluency measures, the four temporal variables SR (speech rate), AR (articulation rate), MLR (mean length of run) and PTR (phonation time ratio) were adopted to measure the EFL learners' spoken English fluency, PRAAT was utilized to extract the temporal indices. And Formula/Run Ratio (FRR) was adopted from previous studies to measure the EFL learners' use of two-word formulaic sequences and three-word lexical bundles.

As for the distribution of the EFL learners' English formulaic language use in their speaking performance, those two types of formulaic language, namely, two-word formulaic sequences and three-word lexical bundles have slightly different distributions across the EFL learners' speech samples. And three-word lexical bundles have been found to be more frequently used by such EFL learners. And this also confirms the study of Vilkaitė (2016), arguing that lexical bundles are the most common formulaic language in the corpus. And the four categories of formulaic language also have different distributions, with three-word lexical bundles being the most frequently used category, followed by two-word

collocations and two-word idiomatic phrases. However, two-word phrasal verbs have been found to be the least frequently used formulaic language category. Such a finding also supports the study of Vilkaitė (2016), which reports that the frequency order of the four formulaic language categories (lexical bundles, collocations, phrasal verbs and idiomatic phrases) in the corpus is lexical bundles, collocations, idiomatic phrases and phrasal verbs.

Based on the results, it might be safe to say that there is indeed a noteworthy relationship between EFL learners' use of English formulaic language and their spoken English fluency, suggesting that EFL learners' spoken English fluency can be facilitated by their use of formulaic language. Therefore, to promote the development of EFL learners' spoken English fluency, they should be encouraged to increase their use of formulaic language such as two-word formulaic sequences and three-word lexical bundles. And this finding also confirms the study of Wood (2015) and Thomson (2017) which indicate that there might be a link between spoken fluency and formulaic language use.

### 5.1 Pedagogical Implications

The findings of this study could lend us much help to further discuss about what formulaic language can be taught in EFL spoken English classes. Specifically speaking, in EFL teaching, compared with two-word collocations and two-word idiomatic phrases, two-word phrasal verbs should be given much more attention since they have been found to be the weakest point of the EFL learners. And the result echoes the study of Hook (2002) arguing that two-word verbs are especially difficult for learners of English as a second language. Therefore, the phrasal verb knowledge of EFL learners should be promoted. Just as what Al Nasarat (2018, p.124) suggested that "we should get learners exposed more to the phrasal verbs by providing more material covering phrasal verbs context and meaning."

In addition, both two-word lexical phrases and three-word lexical bundles have been found to be significant factors that may facilitate the EFL learners' spoken fluency, therefore, EFL learners could be encouraged to practice producing more two-word lexical phrases and three-word lexical bundles in their English speaking to improve their spoken English fluency.

### 5.2 Limitations & Future Research





To be honest, this preliminary study has some limitations, the first being the limited monologue speech sample size. Because, on one hand, the limited English speech samples of the EFL learners may not fully show their actual temporal features of the EFL learners, therefore, for future research, more participants are to be recruited and more speech sample are to be collected. On the other hand, the speech samples are just monologues of the EFL learners, which might be a little unnatural sometimes with regard to their daily conversations. Therefore, the temporal features captured in this study might be slightly different from that of the participants' daily speech performance. For future research, speech samples are suggested to include the conversation of participants. The second limitation is that this pilot study investigated the relationship between EFL learners spoken English fluency and their use of two-word formulaic sequences and three -word lexical bundles, we didn't address the accuracy and appropriacy of their formulaic language use and that might be a good topic for future research.

### References

- Al Nasarat, S., A. (2018). The dilemma of learning phrasal verbs among EFL learners. *Advances in Language and Literary Studies*, 9(2), 119-125.
- Alarbai, J. et al. (2016). *The Importance of Cultural Context in Teaching Formulaic Sequences to ESL/EFL Students*, ProQuest Dissertations and Theses.
- Biber, D. et al. (1999). *Longman Grammar of Spoken and Written English*. Foreign Language Teaching and Research Press
- Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, 14.3, 275-311
- Buerki, Andreas. (2016). Formulaic sequences: a drop in the ocean of constructions or something more significant? *European Journal of English Studies*, 20(1),15-34
- Christian, Jones. (2015). In defence of teaching and acquiring formulaic sequences. *ELT*, 69(3),319-322
- Coxhead, A. (2018). Replication research in pedagogical approaches to formulaic sequences: Jones & Haywood (2004) and Alali & Schmitt (2012). *Language Teaching*,51(1), 113-123
- Craig Lambert, Judit Kormos. (2014) Complexity, Accuracy, and Fluency in Task-based L2 Research: Toward More Developmentally Based Measures of Second Language Acquisition, *Applied Linguistics*, 35(5),607-614
- Darwin, C. M., & Gray, L. S. (1999). Going after the phrasal verb: An alternative approach to classification. *TESOL Quarterly*, 33(1), 65-83.
- DeKeyser, R. (2012). COGNITIVE BASES OF SECOND LANGUAGE FLUENCY. Normal Segalowitz. London: Routledge, 2010. *Studies in Second Language Acquisition*, 34(1), 158-160
- De Jong, N. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, 41(2), 385-39
- De Jong, C. A. M., & Perfetti, C. A. (2011). Fluency training in the ESL classroom: An experimental study of fluency development and proceduralization. *Language Learning*, 61(2), 533-568.
- Doucet, Joelle Lise Marie. (2014). *Formulaic Language in the Interactions of Children with Autism Spectrum Disorder: A Mixed Methods Multiple Case Study* MA of Applied Linguistics and Discourse Studies
- Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text*, 20(1), 29-62.
- Firth, J. R. (1951). Modes of meaning. In J. Firth (Ed.) *Essays and studies* (pp. 118-149). London: Oxford University Press
- Firth, J. R. (Ed.) (1957). *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press
- Gyllstad, Henrik & Schmitt, Norbert. (2019). Testing formulaic language. Chapter of Siyanova-Chanturia, Anna & Pellicer-Sánchez, Ana. *Understanding Formulaic Language: A Second Language Acquisition Perspective*, Routledge Taylor & Francis Group
- Hatami, S. (2015). Teaching Formulaic Sequences in the ESL Classroom. *TESOL Journal*, 6(1), 112-129.
- Hook, J.N (2002). *Two-word Verbs in English*. Heinle & Heinle
- Ho, Robert. (2014). *Handbook of Univariate and Multivariate Data Analysis with SPSS*. Tylor and Francis Group
- Huang, H. (2012). *The effectiveness of formulaic sequence instruction on Taiwanese EFL learners' oral fluency* (Order No. 3542715). Available from ProQuest Dissertations & Theses Global. (1152525941)
- Lennon. P. (1990) Investigating Fluency in EFL: A Quantitative Approach. *Language Learning*, 40(3), p387-417
- Le-Thi, Duyen, Rodgers, Michael P. H., & Pellicer-Sánchez, Ana. (2017). Teaching Formulaic Sequences in an English-Language Class: The Effects of Explicit Instruction versus Coursebook Instruction. *TESL Canada Journal*, 34(3),111-139.
- Liu, D. (2003). The most frequently used spoken American English idioms: A corpus

- analysis and its implications. *TESOL Quarterly*, 37(4), 671-700
- Liu, D. (2011). The most frequently used English phrasal verbs in American and British English: A multicorpus examination. *TESOL Quarterly*, 45(4), 661-688
- Ludivine, Crible. (2018). *Discourse Markers and (Dis)fluency- Forms and functions across languages and registers*. John Benjamins Publishing Company
- Macis, Marijana and Schmitt, Norbert. (2017). *Language Teaching Research*, 21(3), 321-340
- Mandal, S. et al (2017) Application of canonical correlation analysis to a sensitivity study of constitutive model parameter fitting. *Materials and Design* 132,30-43
- Marinez, R. & Schmitt, N. (2012). A phrasal expression list. *Applied Linguistics*, 33(3), 299-320
- McGuire, Michael, & Larson-Hall, Jenifer. (2017). Teaching Formulaic Sequences in the Classroom: Effects on Spoken Fluency. *TESL Canada Journal*, 34(3), 1-25.
- Marinez, R. & Schmitt, N. (2012). A phrasal expression list. *Applied Linguistics*, 33(3), 299-320
- Nattinger, James. R. (1992) and De Carrico, Jeanette S *Lexical Phrases and Language Teaching*. Oxford University Press.
- Oh, Sehee. (2017). The effects of teaching formulaic language on EFL vocabulary learning. *STEM Journal*, 18(3), 29-49.
- Pawley, Andrew & H. Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J.C. Richards & R. W. Schmidts(Eds.), *Language and Communication*, 197-227. London and New York, NY: Longman
- Quan, T. (2016). *Becoming fluent abroad: Examining the intersections of oral fluency, formulaic language, and learner investment in study abroad* (Order No. 10182798). Available from ProQuest Dissertations & Theses Global; Social Science Premium Collection. (1850915390).
- Rezvani Kalajahi, S., Neufeld, S. & Nadzimah Abdullah, A. (2017). The discourse connector list: a multi-genre cross-cultural corpus analysis. *Text & Talk*, 37(3), 283-310
- Russell, Thurstan Steven (2017). Finding the formula: formulaic language use in Hong Kong primary school English textbooks. University of Birmingham. M.A.
- Schmitt, N. (2000). Key concepts in ELT: Lexical chunks. *English Language Teaching Journal*, 54(4), 400-401.
- Segalowitz, N. (2010). *The cognitive bases of second language fluency*. New York: Routledge.
- Shim, Jaewoo, Lee, Heechul & Jin, Seunghee. (2016). A Study on the Correlation between Self-Efficacy Beliefs and Strategies in English Reading. *The Linguistic Association of Korea Journal*, 24(2). 101-125
- Shin, Dongkwang; Nation, Paul. 2008 *ELT Journal*, 62(4), 339-348
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied linguistics*, 31(4), 487-512
- Siyanova-Chanturia, Anna (2015) *Corpus Linguistics and Ling. Theory*, 11(2), 285-301
- Siyanova-Chanturia, Anna. et al. (2017). Representation and processing of multi-word expressions in the brain. *Brain and language*, 175, 111-122
- Skehan, Peter. (2011). Book and Software Reviews. *Canadian Modern Language Review*, 67(2), 269-271
- Taylor, John R. (2012) *The Mental Corpus: How language is represented in the mind*. Oxford Scholarship Online, 10.1093/acprof:oso/9780199290802.001.0001
- Thomson, Haidee (2017) Building Speaking Fluency with Multiword Expressions. *TESL Canada Journal*, 34(3), 26-53
- Thomson, R. I. (2015) Fluency. In Reed, M. & Levis, J. *The Handbook of Pronunciation*. (209-226). Hoboken, New Jersey: Wiley.
- Underhill, N. (1987). *Testing Spoken Language a handbook of oral testing techniques*. Cambridge University Press
- Van Lancker Sidtis, Diana & Seung-yun Yang (2017) Formulaic language performance in left- and right-hemisphere damaged patients: structured testing, *Aphasiology*, 31(1), 82-99
- Van Lancker Sidtis, Diana & Sidtis, John J. (2018) Cortical-subcortical production of formulaic language: A review of linguistic, brain disorder, and functional imaging studies leading to a production model. *Brain and Cognition*, 126, 53-64
- Vilkaitė, Laura. (2016). Formulaic language is not all the same: comparing the frequency of idiomatic phrases, collocations, lexical bundles, and phrasal verbs *Taikomoji kalbotyra* (8), www.taikomojikalbotyra.lt
- Wood, D. (2006). Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review*, 63 (1), 13-33.
- Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence, and classroom applications*. London: Continuum.
- Wood, D. (2015). *Fundamentals of formulaic language: An introduction*. London/New York: Bloomsbury.



- Weinert, R. The role of formulaic language in second language acquisition: A review, *Applied Linguistics*, 16(2), 180-205.
- Wray, A. (2000), Formulaic Sequences in Second Language Teaching: Principle and Practice, *Applied Linguistics*, 21(14), 463-489
- Wray, A. (2002). Formulaic language and the lexicon. Cambridge: Cambridge University Press.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press
- Wray, A. (2012). What Do We (Think We) Know About Formulaic Languages? An Evaluation of the Current State of Play. *Annual Review of Applied Linguistics*, 32, 231-254
- Wray, A. (2013). Formulaic language. *Language Teaching*, 46, 316-334

---

<sup>i</sup> <https://www.lex tutor.ca/>

<sup>ii</sup>

<https://www.collinsdictionary.com/dictionary/english/collocation>

<sup>iii</sup> <http://www.freecollocation.com/>

<sup>iv</sup> <http://praascripting.lingphon.net/>