

# On the Phraseology of Chinese College Student Oral English in Public Speaking

Yiming Zhang<sup>1</sup>

<sup>1</sup> Beijing Foreign Studies University, PR China

Correspondence: Yiming Zhang, Beijing Foreign Studies University, PR China.

doi:10.56397/JLCS.2023.12.02

## Abstract

Prior investigations of phraseology have systematically researched the teaching and learning of phraseology in EFL setting. Nevertheless, little research has been done specifically on the phraseology of Chinese college student oral English in public speaking. This study sets out to study the formal and functional features of phraseology in a self-built corpus containing 300 speeches delivered by Chinese college students by comparing them against those of native speakers. It has been found that the overall use of phraseology by Chinese college students bears a resemblance to that by native speakers across the four structural types of chunks, including polywords, phrasal constraints, institutionalized expressions and sentence builders. On the other hand, Chinese college students show some unique features in the use of phraseology, including the lack of diversity and the over-reliance on some chunks, etc. The findings of this study also provide significant pedagogical implications for teaching spoken English, especially English public speaking.

**Keywords:** phraseology, Chinese college students, oral English, English public speaking, corpus study

## 1. Introduction

Phraseology, also known as lexico-grammatical units, has been a subject of linguistic research over an extended period, which provides a promising new direction for language teaching (Nattinger & Decarrico, 1992). By definition, phraseology is multi-word lexical phenomena or form/function composites that occur more frequently and bear more idiomatical meanings than expressions constructed by linguistic conventions or rules.

The research of phraseology can be divided into three categories: the structures and functions of phraseology, practical value of it, and the comparison of semantic and collocational

features of the use of phraseology across various discourses or groups of language users (especially between English L1 users and their L2 counterparts). Those studies which mainly focus on examining the first two issues include Peter (1983), Pawley and Syder (1983), Bible et al. (2004), Hyland (2008a). These studies have discussed the relationship between the form and function of phraseology with a limited number of lexical phrases and a myriad of them focus on whether phraseology should fall into the category of competence, performance, pragmatics, or some integration of these.

As for the third focus point, researchers in corpus linguistics have been uncovering more linguistic features of lexical phrases by

conducting contrastive study in the past few decades. With the help of computers, researchers can handle large quantities of language data. This makes it possible to explore with more accuracy the pedagogical value of phraseology, as well as semantic and collocational differences of the use of phraseology. For example, pedagogical value of phraseology has been investigated by Boers (2005), Ding (2005), Biber (2009), Ellis (2012), among others. Studies conducted by Romer (2009), Chen and Baker (2010), Ädel & Erman (2012) reveal some significant semantic and collocational differences of the use of FL between L1 and L2 learners. Research done by Cortes (2002), Cortes (2004), Hyland (2008b) compared the differences between novice and professional writers. These corpus-driven studies reveal that phraseology can play a pivotal role in both language acquisition and language use.

So far, however, few studies have been conducted systematically as to the formal and functional features of the phraseology of Chinese college student oral English in public speaking based on relatively large quantities of data. This study aims to bridge the gap. The following two issues, which merit methodical and meticulous investigation, will be discussed in this paper: 1) In terms of formal and functional features, what are the main hallmarks of Chinese college students' use of phraseology in English public speaking? 2) How much do these hallmarks resemble or differ from those of native speakers? Answers to these two questions should contribute to a more comprehensive understanding of phraseology and lend considerable support to the teaching of English public speeches oriented towards English L2 learners.

This paper sets out to examine and illustrate the phraseological features of Chinese college student oral English in public speaking, predicated on the evidence collected from two self-built corpora. The objective of this study can be twofold. First and foremost, I would like to provide a general overview of the formal and functional characteristics of Chinese college students' usage of phraseology in English public speaking across four types of chunks. In addition, I wish to compare the use of phraseology by Chinese college students and native speakers in terms of several essential formal and functional elements.

In the following sections, I shall first offer a brief review of the previous studies on phraseology. Then, I shall introduce the corpora used in this study and offer a brief overview of the methodology. After that, I will demonstrate the overall distribution of the phraseologies in the corpora. Furthermore, I will sort out in detail the formal and functional features of chunks in the key categories of phraseology by discussing the quantitative and qualitative similarities and differences between the phraseology used by Chinese college students and that by native speakers. Finally, I will summarize the study's principal findings and provide their implications, in particular, pedagogical implications.

## 2. Literature Review

There has been a long history of phraseology research in applied linguistics, dating back to Firth (1951), who puts emphasis on the term "collocation" and argues that "you shall know a word by the company it keeps." Nattinger & DeCarrico (1992) first put forward a systematic way to classify phraseology according to its structures, identifying four types of phraseology, i.e., polywords, institutionalized expressions, phrasal constraints, and sentence builders. Biber et al. (1999) defined phraseology as "word sequences that occur more than ten times per million words and are distributed in more than five texts."

Over the past decades, some linguists have investigated the structures and functions of phraseology and the practical value of it, while other studies have been largely concerned with the comparison of semantic and collocational features of the use of phraseology across various discourses or language users. There has been a large body of literature on the investigation of phraseology from a functional perspective. Based on the different functions of phraseology, Biber et al. (2004) categorized phraseology into stance bundles, discourse organizers, referential expressions, and special conversational bundles. It is also observed that the use of these types of phraseology are significantly different in spoken and written texts. For instance, more stance bundles are observed to be used in speaking, while more referential expressions are used in academic writing. Hyland (2008a) also proposed a method to classify phraseology in academic papers from a functional perspective, dividing phraseology into research-oriented, text-oriented and participant-oriented types.

Findings of Weinert (1995), Wood (2002), Boers (2005), Ding (2005), Biber (2009), and Ellis (2012) show that there should be a direct correlation between phraseology and language proficiency. For instance, Boers (2005) performed a computer-aided experiment to analyze the use of alliteration and concluded that salient phonological patterning (e.g., alliteration) can facilitate learning and memorizing phraseology, and further increase the learners' language proficiency. Besides, Ding (2005) explored the connection between a student's English level and the frequency of the use of phraseology. It is found that phraseology can help students to speak and write English in a more idiomatic and accurate way. These studies have been largely concerned with the ways to facilitate phraseology learning and have also provided evidence for the positive effects of phraseology acquisition upon language proficiency.

More recently, studies have observed that the systematic differences of phraseology not only exist in different discourses, but also in different language users. A growing body of literature has been focusing on the semantic and collocational differences of phraseology across various groups of language learners. Most studies have been conducted (e.g., Romer, 2009; Chen & Baker, 2010; Ädel & Erman, 2012) on the differences of the use of phraseology between L1 and L2 learners. For instance, Romer (2009) noted that in advanced-level writing, L1 and L2 learners demonstrate similar ability in using phraseology because experience and expertise are more important than nativeness in academic writing. Other studies have been largely concerned with the comparison of the use of phraseology between novice and expert writers in academic writing (e.g., Cortes, 2002; Cortes, 2004; Hyland, 2008b). For instance, Cortes (2002) investigated bundles in compositions written by college freshmen and found that the bundles employed by these novice writers are functionally different

from those in published academic prose. Hyland (2008b) noticed that novice writers, including postgraduate students, tend to use more formulaic expressions than expert writers to display their competence.

The literature mentioned above have systematically documented theory-driven and corpus-driven studies on phraseology in terms of the structures and functions, the practical value, the comparison of the semantic and collocational features across various groups, etc. One potential limitation, it seems, is that they rarely pay attention to the phraseology of EFL learners in public speaking. To bridge this gap, the current study adopts a corpus approach to analyze Chinese college students' use of phraseology in English public speaking. I wish to explore what phraseological features characterize Chinese college students' use of language in English public speaking, and to what extent those features resemble, or differ from, those of native speakers.

### 3. Methodology

The present study uses data from a self-built Chinese Learner English Public Speaking Corpus (CLEPSC). It is a 130,800-word corpus which contains 300 speeches delivered by contestants in the two influential English public speaking contests in China, (i.e., "FLTRP Cup" National English Public Speaking Contest and "21<sup>st</sup> Century Cup" National English Public Speaking Competition). Raw materials for CLEPSC are the impromptu speeches from 2010 to 2020 in the two contests mentioned above.

The selected impromptu speeches were transcribed and annotated, excluding fillers, pauses, repeats, and conversational turns, as well as phonological features such as various forms of mispronunciation, misplaced stresses, non-verbal sounds, and indistinct noises. The overall information of the corpus is shown in Table 1.

**Table 1.**

|                    | NUMBERS |                | NUMBERS |
|--------------------|---------|----------------|---------|
| Impromptu speeches | 300     | 4-letter words | 25,703  |
| Speech topics      | 27      | 5-letter words | 13,773  |
| Tokens             | 130,800 | 6-letter words | 7,857   |
| Types              | 6,824   | 7-letter words | 7,002   |
| Type/token ratio   | 5.22    | 8-letter words | 4,201   |

|                       |        |                    |       |
|-----------------------|--------|--------------------|-------|
| Std. type/token ratio | 38.23  | 9-letter words     | 3,322 |
| Av. word length       | 4.03   | 10-letter words    | 2,205 |
| Sent. length          | 14.45  | 11-letter words    | 965   |
| Std. sent. length     | 13.81  | 12-letter words    | 583   |
| 1-letter words        | 6,832  | 13-letter words    | 248   |
| 2-letter words        | 29,560 | 14(+)-letter words | 65    |
| 3-letter words        | 28,484 |                    |       |

The reference corpus consists of 55 speeches delivered by native speakers selected from a speech corpus, American Rhetoric (AR). These speeches were carefully selected to reduce the repetition of topics and to make the size of the

reference corpus commensurate with the research one. The token of the corpus amounts to 130800, and the TTR is 5.29. Table 2 demonstrates the overall information of the reference corpus.

**Table 2.**

| NUMBERS                 |         | NUMBERS            |        |
|-------------------------|---------|--------------------|--------|
| Native-speaker speeches | 55      | 4-letter words     | 25,645 |
| Speech topics           | 14      | 5-letter words     | 13,002 |
| Tokens                  | 130,972 | 6-letter words     | 8,984  |
| Types                   | 6,925   | 7-letter words     | 7,657  |
| Type/token ratio        | 5.29    | 8-letter words     | 5,983  |
| Std. type/token ratio   | 39.04   | 9-letter words     | 4,032  |
| Av. word length         | 4.22    | 10-letter words    | 2,543  |
| Sent. length            | 16.01   | 11-letter words    | 1,264  |
| Std. sent. length       | 15.18   | 12-letter words    | 620    |
| 1-letter words          | 6,303   | 13-letter words    | 289    |
| 2-letter words          | 28,166  | 14(+)-letter words | 96     |
| 3-letter words          | 26,388  |                    |        |

The research method employed in this study is as follows. Predicated on previous research (e.g., Biber, 1999; Wray, 2002) and the aim of this study, I temporarily define phraseology as “continuous or discontinuous word combinations consisting of two to six words, with relatively complete meanings and high frequencies in CLEPSC”. In the first step, I used the N-Gram model in AntConc to retrieve phraseologies from CLEPSC. The N-Gram Size was set to be two to six, so that I was able to obtain 2-word, 3-word, 4-word, 5-word, and 6-word chunks. Secondly, I established cut-off frequencies for phraseology of various lengths to exclude the less common chunks, with the cut-off frequencies for 2-word to 6-word chunks being 4, 4, 3, 3 and 3, respectively. The reasons are twofold: 1) Chunks which have a very low occurrence can occur in

the corpus quite by chance; 2) There is a massive amount of data in the corpus, which will be hard to study if not screened out with a set of frequency criteria. In the third step, I manually removed those semantically unacceptable word combinations which were not qualified as phraseology, such as “think the”, “that we can”, “and I believe the”, “all thank you for your”, etc. Fourthly, in the qualitative analyses, I adopted Nattinger & DeCarrico’s (1992) framework to describe and discuss the formal and functional features of the phraseologies, categorizing them into four categories, i.e., polywords, institutionalized expressions, phrasal constraints, and sentence builders and then compared them with native speakers. Four structural criteria are used in the classification of the four types of phraseologies: The first is their length and

grammatical status; the second is whether the phrase has a canonical or non-canonical shape; the third is whether the phrase is variable or fixed; and the fourth is whether the phrase is continuous or discontinuous, i.e., whether it consists of an uninterrupted sequence of words or is interrupted by lexical items separating the phrase into several parts.

#### 4. Results and Discussion

In the following sections, I shall first present the overall data and structural types of phraseology from a general perspective, and then discuss the major functional and formal features of the four structural types of chunks in detail.

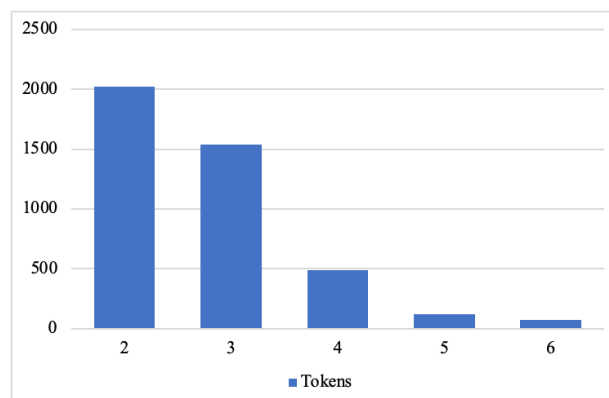
##### 4.1 Overall Statistics and Structural Types of Phraseology

4248 chunks (232 types) were obtained based on manual check. The general distribution of the phraseologies can be seen in Table 2. The frequencies of the phraseologies categorized by lengths is graphically depicted in Figure 1.

**Table 3.** Overall distribution of the phraseology in CLEPSC

| CHUNKS       | TYPES | TOKENS |
|--------------|-------|--------|
| 2-word chunk | 92    | 2024   |
| 3-word chunk | 69    | 1541   |
| 4-word chunk | 46    | 489    |
| 5-word chunk | 14    | 122    |
| 6-word chunk | 11    | 72     |

|       |     |      |
|-------|-----|------|
| Total | 232 | 4248 |
|-------|-----|------|



**Figure 1.** Frequencies of the phraseology ranging from two to six words in CLEPSC

As shown by Table 3 and Figure 1, the overall distribution demonstrate that 2-word chunks account for the greatest proportion of all chunks, occurring 2024 times and covering 92 types. 3-word chunks are the second most frequent, with the number of 4-word chunks slumping to around 30% of that of 3-word chunks. The types of 5-word and 6-word chunks bear a resemblance to each other while dramatically lower than those of the other three types.

The distribution of the four types of phraseologies, i.e., polywords, institutionalized expressions, phrasal constraints, and sentence builders, is demonstrated in Table 4 and Table 5.

**Table 4.** Frequency distribution of the four types of phraseology in CLEPSC

| CLASSIFICATION                | TOKEN | TYPE | TYPE/TOKEN RATIO | PERCENTAGE |
|-------------------------------|-------|------|------------------|------------|
| Polywords                     | 385   | 41   | 10.65            | 9.06       |
| Institutionalized expressions | 712   | 17   | 2.39             | 16.76      |
| Phrasal constraints           | 993   | 76   | 7.65             | 23.38      |
| Sentence builders             | 2158  | 98   | 4.54             | 50.80      |
| Total                         | 4248  | 232  | 5.46             | 100.00     |

**Table 5.** Frequency distribution of the four types of phraseology in the reference corpus

| CLASSIFICATION                | TOKEN | TYPE | TYPE/TOKEN RATIO | PERCENTAGE |
|-------------------------------|-------|------|------------------|------------|
| Polywords                     | 668   | 68   | 10.18            | 15.82      |
| Institutionalized expressions | 236   | 32   | 13.56            | 5.59       |



|                     |      |     |      |        |
|---------------------|------|-----|------|--------|
| Phrasal constraints | 1362 | 121 | 8.88 | 32.25  |
| Sentence builders   | 1957 | 71  | 3.63 | 46.34  |
| Total               | 4223 | 292 | 6.75 | 100.00 |

As is revealed in Table 4, sentence builders, with their occurrences of 2,158, are the most frequently occurring chunks, accounting for 50.80% of the total. Phrasal constraints are the second most frequently used chunks, but their tokens are not even half of sentence builders', accounting for a dramatically smaller percentage of the total. Polywords, on the other hand, occur only 385 times, accounting for the lowest percentage of the total. The uneven distribution of the four types of phraseology suggests that Chinese college students in English public speaking have a predilection for particular types of phraseology.

The overall distribution of data in CLEPSC bears a resemblance to that in the reference corpus. The total frequency in the reference corpus is 4323, similar to 4248 in CLEPSC. Besides, sentence builders have the highest frequency of occurrence and percentage, both in CLEPSC (2158, 50.80%) and in the reference corpus (1957, 46.34%). The similarity in the use of sentence builders seems to indicate that both EFL learners and native speakers prefer to use chunks which are larger and contain more information than discrete items, e.g., sentence builders, and their size increases as we become more familiar with remembered material, permitting us to store and recall more information (Simon, 1974). Speech production is characterized by this sort of chunks to a surprising degree. It is the ability to use lexical phrases, especially long phrases like sentence builders that helps us speak more fluently. Sentence builders allow language users to focus more on the larger structure of the discourse, rather than narrowly on individual words with which they build sentences from scratch.

Upon further study, however, it has also been noticed that there are distinct differences across the two groups. For example, Chinese college students cannot use as many polywords as native speakers (385 versus 668). Institutionalized expressions used by Chinese college students are also substantially lower in number than those by their counterparts. As a result, Chinese college students may sound less natural and less fluent than native speakers in

English public speaking. This will be returned to in more detail in the following sections.

#### 4.2 Major Functional and Formal Features of Sentence Builders

Sentence builders are incomplete sentence-level lexico-grammatical sequences for building sentences. They are frames or skeletons which can be filled in with appropriate phrases or clauses. Functionally speaking, a great many of them are highly conventionalized formulae associated with indirect speech acts. Due to the demands of impromptu speaking which often requires speakers to make sentences within a very limited amount of time, speakers usually turn to this kind of ready-made lexical phrases which contain slots for parameters or arguments for expression of entire ideas. The following description and discussion illustrate the major pragmatic functions of sentence builders in the research and reference corpora.

Upon close observation of the data collected from the corpora, we find both similarities and differences between Chinese college students and native speakers. In functional terms, these sentence builders can further sub-divided into 9 categories, i.e., assertion, relators, summarizer, evaluator, topic marker, qualifier, request, topic shifter, and comparator (Nattinger & Decarrico, 1992). Among them, 3 categories are unique in the research corpus, including relators, request chunks, and comparators. Table 6 gives a sample of the most frequently occurring chunks for each functional category.

**Table 6.** Major functions of sentence builders in CLEPSC

| FUNCTIONS    | INSTANCES                    | FREQ. |
|--------------|------------------------------|-------|
| Assertion    | I think/believe (that) X     | 62    |
| Relators     | not only X, but also Y       | 26    |
| Summarizer   | my point is that X           | 22    |
| Evaluator    | It's a good idea to do X     | 15    |
| Topic marker | my topic (today) is (that) X | 33    |
| Qualifier    | it's only in X that Y        | 12    |

|               |                                      |    |
|---------------|--------------------------------------|----|
| Request       | I want/would like (somebody) to do X | 6  |
| Topic shifter | that reminds me of X                 | 11 |
| Comparator    | the -er X, the -er Y                 | 5  |

Similarities can be found as regards the syntactic frameworks and actual utterances. From the perspective of phraseology, sentence builders presented in Table 4 are relatively simple syntactic frameworks. Nevertheless, they are exceedingly flexible in that the slots and fillers enable a myriad of different actual utterances, both in CLEPSC and in the reference corpus. One case in point might be the sentence builder below which functions as request chunk: *I want/would like (somebody) to do X*, with noun phrase or gerund expansion being filled in as X. A host of actual utterances based on this form can be found in CLEPSC and the reference corpus, such as *I want everybody here to join me on looking for that part*, *I want you to think about it*, *I'd like to invite you to look at it from a different perspective*, etc. In other words, for any particular function, sentence builders used by Chinese college students and native speakers are rife with paradigmatic slot/filler variations, albeit usually with some simple syntagmatic frames. Furthermore, it seems that the two groups tend to employ similar sentence builders to achieve specific speech act functions under specific contexts or circumstances. For instance, relators are used to introduce the speaker's opinions; request phraseology is used to appeal to the audience for action; comparator phraseology is used for expressing causal or progressive relationship between the two subjects the speaker intends to compare. In actuality, the reason why these three unique functional categories of phraseology occur in the research corpus may well be that the directions given in speech contests bear a resemblance to the tasks native speakers need to fulfill in public speaking, such as expressing personal views, calling for action, and comparing different opinions, etc.

Although the two sets of data demonstrate similarity in terms of the overall quantitative distributions, many types of sentence builders used by Chinese college students are considerably different from those by native speakers. In the first place, Chinese college students' use of sentence builders reflect their non-nativeness and unnaturalness as compared to the way native speakers deploy phraseology

in their speeches. I have found, upon further study, that phraseology used by Chinese college students is small in type, albeit large in number. For example, in the research corpus, frequently occurring sentence builders of assertion are *I think (that) X*, *I believe (that) X*, with occurrences of 43 and 19; percentages of 1.99% and 0.88% in terms of all the sentence builders, respectively. In comparison, native speakers can use a wider scope of sentence builders to make assertion. *I think (that) X* only occurs 31 times and accounts for 1.58% of sentence builders in the reference corpus. *I believe (that) X* has a frequency of mere 12 and accounts for 0.61%. However, more types of phraseology expressing assertion are found in the reference corpus, e.g., *I argue (that) X*, *I (would) say (that) X*, *I claim (that) X*, *I hold (that) X*, etc. This finding suggests that compared with native speakers, Chinese college students are likely to rely more on a small number of chunks in speeches, whereas unable to cover more varieties. To the native ear, too much recurrence of few types of phraseology may seem repetitive and unnatural.

Furthermore, it should be noted that there are a few disparities in the use of topic markers across the two groups. Quantitatively speaking, Chinese college students employed much more instances of topic markers than native speakers. It is observed that the topic marker *my topic (today) is (that) X*, with a token of 33, are the second most frequently used sentence builders, accounting for 0.78% of the total chunks in the research corpus, while they only occur 5 times and account for 0.12% in the case of the reference corpus. This seems to indicate that Chinese college students are, comparatively speaking, more dependent on the prefabricated pattern *my topic (today) is (that) X*, among others, to introduce their topics, rather than compose their own topic-introducing sentences from scratch using grammatical rules, the latter of which requires a speaker to be highly proficient at the language.

Last but not the least, we also notice that cultural differences play a large part in forming those formal and functional differences, hence the necessity to analyze them from a socio-pragmatics perspective (Leech, 1983). For example, Chinese college students have a predilection for the evaluator *It's a good idea to do X*. In semantical terms, to express the same meaning, native speakers tend to deploy another evaluator *I'm a great believer in X*, e.g. *I'm a great*

*believer in putting money away for a rainy day, I'm a great believer in the power of change.* The word *believer*, often defined as a person who believes in the existence of a particular God or religion, bears with itself some religious implications. Most native speakers are Christians, and thus *believer* is a term they are quite familiar with. But Chinese college students, more often than not, know little about this cultural meaning behind the literal meaning of the chunk. Therefore, when they need to evaluate something, the first chunk they can think of tend to be the former rather than the latter one. To summarize, the religious view has a profound impact upon the language use and has resulted in the English chunk *I'm a great believer in X*, which bears a semantical resemblance to another chunk *It's a good idea to do X*, which, nevertheless, are more frequently used by Chinese college students in English public speaking. Presented below are two extracts taken from the speeches in the research and reference corpus.

(1) <sp1> It's a good idea to step out of our comfort zone, little by little, not too much at once, and remember that our achievements came at a price. </sp1>

<sp2> I'm a great believer in hope – hope in the face of difficulty; hope in the face of uncertainty; the audacity of hope! </sp2>

As is shown in the extracts above, the speaker in “sp1”, a Chinese college student, used the sentence builder *It's a good idea to do X* to demonstrate his evaluation on the idea “stepping out of our comfort zone”, i.e., avoiding complacency and facing challenges bravely. The native speaker in “sp2” talked about the value of hope by using the evaluator *I'm a great believer in X*, calling on everyone not to lose hope.

To sum up, there are both similarities and differences in terms of formal and functional properties of sentence builders in CLEPSC and the reference corpus. Firstly, syntactic frameworks and actual utterances of sentence builders in CLEPSC seem to resemble those of their native speaker counterparts, based on the data showing a large proportion of sentence builders with simple syntactic structures but sophisticated actual utterances. The functions of sentence builders realized in CLEPSC also seem to resemble those in the reference corpus. Secondly, some sentence builders in CLEPSC appear to have formal features that are

significantly different from those of their counterparts in the reference corpus, insofar as the remarkable differences across the two groups in the proportions of assertion phrases, topic markers, etc. Thirdly, some socio-pragmatic differences of the use of phraseology have been found, shedding light on the possible impact culture can have on the use of phraseology.

#### 4.3 Major Functional and Formal Features of Polywords

Polywords are short phrases that function similarly to individual lexical items. They generally have three characteristics: (1) They can be both canonical and non-canonical. (2) They don't allow for any variation. (3) They are continuous. Polywords are associated with a myriad of functions. Common ones that can be observed in English public speeches include conveying speaker qualification for the current topic, linking one topic to another, shifting topics, summarizing, etc. The data of polywords in CLEPSC is shown in Table 7.

**Table 7.** Major functions of polywords in CLEPSC

| FUNCTIONS        | INSTANCES          | FREQ. |
|------------------|--------------------|-------|
| Qualifier        | in part            | 5     |
| Summarizer       | in conclusion      | 6     |
| Topic shifter    | by the way         | 8     |
| Agreement marker | I'll say           | 18    |
| Fluency device   | at any rate        | 13    |
| Relator          | for that reason    | 9     |
| Evaluator        | generally speaking | 11    |
| Clarifier        | you know           | 40    |

Both similarities and differences can be found across the two groups. The most striking similarity lies in the use of fluency devices. The frequency of fluency devices account for 13.87% of the total frequency of polywords in CLEPSC, resembling the percentage of 12.99% in the reference corpus. The types of fluency devices in CLEPSC are also similar to those in the reference corpus. In functional terms, fluency devices lengthen the sentence to give the speaker more time to plan for what they intend to express next, and thus promotes fluency. The following few



sentences in CLEPSC (“sp1”, “sp2”) and in the reference corpus (“sp3”, “sp4”) can serve as good examples:

(2) <sp1> At any rate, we need to keep finding something to fight for. </sp1>

<sp2> At any rate, that is what we are going to try to do. </sp2>

<sp3> ...and seventy-five out of every hundred want to take care of their own children, which, at any rate, makes full-time job impossible. </sp3>

<sp4> At any rate, as a great nation, we can do this together to make his goal of peace come true. </sp4>

In the abovementioned instances, the deletion of *at any rate* barely changes the semantic function of each sentence. However, wiping it out defies the speaker more time to gather thoughts for the next part. It would be unreasonable to condemn such fluency devices as linguistic crutches or as verbose, empty filler (Nattinger & Decarrico, 1992). One of the crucial reasons, perhaps, is that they serve an extremely important function of promoting fluency. It seems that both Chinese college students and native speakers use fluency devices abundantly to gain more time to plan subsequent discourse and to promote fluency.

The disparities between the two groups can be threefold. In the first place, the overall use of polywords is different. Standard frequency has been calculated to compare the overall use of polywords across the two groups. Data show that polywords occur 2200 more times every 1,000,000 words in CLEPSC than in the reference corpus. This indicates that there is a remarkable difference between Chinese college students and their counterparts in terms of the total frequency of polywords. From a formal perspective, polywords used by Chinese college students are mainly canonical chunks, while noncanonical ones barely occur in CLEPSC. Nevertheless, both canonical and noncanonical chunks can be found in the reference corpus. The noncanonical ones which are unique to the reference corpus include *as it were* (exemplifier), *so far so good* (approval marker), *once and for all* (summarizer), *in essence* (summarizer), *by and large* (qualifier), etc.

Secondly, in terms of functional subtypes, Chinese college students show a lack of variety and range when they use each category of polywords compared with native speakers. For

instance, only 4 types of summarizers have been observed in CLEPSC, i.e., *in conclusion* (6 times), *in summary* (4 times), *to summarize* (3 times), *in a word* (3 times). However, native speakers employ a wider range of summarizers, with types amounting to at least 7 in total. These summarizers include *in conclusion* (5 times), *all in all* (5 times), *in brief* (5 times), *as I have said* (5 times), *in short* (4 times), *to wrap up* (4 times), *in a nutshell* (3 times). Similar results can also be observed in terms of relators (4 versus 6 types) and clarifiers (3 types versus 8 types), among others. This suggests that native speakers utilize polywords in a broader range and are capable of substituting one polyword with a large number of others that convey the same meaning in order to balance the frequency of polywords used. It is clear that Chinese college students depend too heavily on particular polywords in their speech, making them seem less natural or idiomatic, whereas native speakers use a variety of alternative phrases to convey the same idea. In addition, I also discovered that Chinese students made more errors in polywords than native speakers. For example, Chinese college students use *in the campus* up to 11 times in total, although the right term is *on the campus*. Pedagogically, Chinese college students rely on some specific types of polywords as they are chunks learned in the early period of their English studying. Besides, it should also be noticed that in Table 7, there are no disagreement markers such as *not on your life*, *hold your horses*, etc., which might decrease the pragmatic quality of discourse.

#### 4.4 Major Functional and Formal Features of Institutionalized Expressions

Institutionalized expressions generally possess the following four formal features: (1) Institutionalized expressions are sentence-length lexical phrases that function as independent utterances. (2) The majority of them are canonical. (3) They are invariable. (4) The majority of them are continuous. They are thus extremely similar to polywords in the last two regards. Institutionalized expressions include proverbs, aphorisms, formulas for social interaction, and many other expressions that a speaker may find useful to store as language units. They are widely employed for citation and allusion. Some of them may be common phrases, especially phatic language, used by practically everyone in the speech community, such as *how are you*, while others may be more idiosyncratic

phrases to convey an idea effectively and efficiently, such as *give me a break*, *have a nice day*, etc. They are mostly continuous. However, discontinuous institutionalized expressions play a significant role in all sorts of texts, speeches in particular, despite their small proportion. Therefore, I will discuss the use of both types of institutionalized expressions in English public speaking. There are 10 major subtypes of institutionalized expressions based on their functions: advice, greeting, closing, parting, narrative framer, disapproval, objection, approval, denial, concession. Seven of them are above the cutting-off frequency. Shown in Table 8 are the frequency of these seven types of phraseology and the corresponding samples.

**Table 8.** Major functions of institutionalized expressions in CLEPSC

| FUNCTIONS   | INSTANCES                    | FREQ. |
|-------------|------------------------------|-------|
| Advice      | no pain, no gain             | 4     |
| Greeting    | good morning/afternoon       | 95    |
| Closing     | thank you                    | 281   |
| Disapproval | beat around the bush         | 5     |
| Objection   | to make matters/things worse | 7     |
| Approval    | silver lining                | 8     |
| Concession  | even so                      | 13    |

The overall use of institutionalized expressions in CLEPSC is significantly different from that in the reference corpus. In terms of standard frequency, Chinese students use 3600 more institutionalized expressions every 1,000,000 words than native speakers. I believe, however, that this cannot prove that Chinese students are more capable of employing institutionalized expressions than native speakers but may well indicate that Chinese college students tend to rely on certain institutionalized expressions such as *thank you*, *good morning/afternoon* too much. Chunks like these are highly recurrent in CLEPSC, which is an indicator of the limitedness of Chinese college student English proficiency.

Significant differences across the two groups have also been found concerning the use of some specific chunks, two of which merit detailed discussion, i.e., *thank you* and *good*

*morning/afternoon*. The closing chunk *thank you* is used very often by Chinese college students. It occurs 281 times and accounts for 6.61% of all the chunk-tokens, which is the highest ratio for all the chunks. Nevertheless, in the reference corpus, *thank you* only occurs 39 times and makes up 0.92%. Despite the sharp contrast in frequency and percentage, the reason behind it cannot be simply put as “different predilection for the chunk *thank you*”. The underlying reason of it, it seems, can be expounded as follows. *Thank you* is usually employed to indicate the end of a speech. In this regard, normally this chunk only occurs once in a speech, and usually at the end of it. The sharp contrast in the number of this chunk, therefore, may be contributed to the difference of the total number of speeches between CLEPSC and the reference corpus (300 versus 55). To better illustrate this issue, I calculated the ratio of frequency to speech numbers (frequency/speech numbers) and found that there is no significant difference between these two groups (93.67% versus 70.90%).

As for the use of the greeting chunk *good morning/afternoon*, the same reason will not suffice to explain the stark contrast between the two groups. This chunk merely has a frequency of 4 in the reference corpus while occurs up to 95 times in CLEPSC. Significant difference has been found in both the frequency (95 versus 4) and the ratio of frequency to speech numbers (31.67% versus 7.27%). Given this, it seems to be appropriate and reasonable to point out that Chinese college students are more dependent upon the use of the chunk *good morning/afternoon*. Under most circumstances, they tend to use this sequence to greet the audience and begin their speeches. They seem to regard this as a conventionalized and indispensable part in English public speaking. This feature for beginning a speech is unique to the speech of Chinese college students, while native speakers tend to use *hello* to start their speeches or get straight to the point without greeting the audience. The underlying reason, which may be associated with socio-pragmatics, merits further investigation in future studies.

#### 4.5 Major Functional and Formal Features of Phrasal Constraints

Phrasal constraints bear the following four formal features: (1) Phrasal constraints are phrases of various lengths, ranging from short ones to medium-length ones. (2) They can be

canonical or non-canonical. (3) They are incomplete phrases with one or more slots, with various lexical and phrasal categories (NP, VP, N, V, Adj, Adv, etc.) can be filled in to form variations. (4) The majority of them are continuous. Phrasal constraints, like polywords, are connected with a wide range of functions.

In functional terms, phrasal constraints can be categorized into 14 categories: temporal relator, summarizer, topic shifter, greeting, parting, closing, relator, qualifier, disapproval, comparator, evaluator, exemplifier, cause, desire. These pieces are not grammatically well-formed, nor psychologically significant. They are,

nonetheless, capable of producing whole sentences. Well-structured sequences can be built by combining them with a variety of lexical and phrasal items. The high frequency (993) and percentage (23.38%) of phrasal constraints in CLEPSC suggest that they are a crucial tool for generating speech. Along the lines of Nattinger & Decarrico (1992), in the following examples, I used slots to show the positions filled by paradigmatic replacement while parentheses to indicate optional syntagmatic information that is not a part of the fundamental frame. In Table 9, I presented 8 major functional types and provided one sample for each type.

**Table 9.** Major functions of phrasal constraints in CLEPSC

| FUNCTIONS        | INSTANCES                 | VARIATIONS   | FREQ. |
|------------------|---------------------------|--|-------|
| Temporal relator | a __ ago                  | a year ago, a very long time ago                               | 13    |
| Summarizer       | to __ this up             | to tie this up, to wrap this up                                | 4     |
| Topic shifter    | as I __                   | as I was saying, as I see it                                   | 6     |
| Relator          | __ as well as __          | this one as well as that one, my friends as well as my parents | 52    |
| Qualifier        | as far as __              | as far as I know, as far as I am concerned                     | 20    |
| Comparator       | the __er the __er         | the sooner the better, the more the better                     | 8     |
| Evaluator        | for better or (for) worse | for better or worse  | 5     |
| Exemplifier      | for __                    | for instance, for example                                      | 115   |

The frequency of the total phrasal constraints used by Chinese college students is significantly different from that by native speakers. Native speakers use 2800 more phrasal constraints every 1,000,000 words than Chinese college students. There is a positive correlation between phrasal constraints and speech coherence. In this regard, it seems to be the case that Chinese college students' speeches are less coherent than those delivered by native speakers. On the other hand, although the types of phrasal constraints used by Chinese college students (76) are also fewer than those of native speakers (121), there is no significant difference across the two groups.

In terms of the 8 major functional types listed in Table 9, there are conspicuous discrepancies across the two groups. Many chunks, such as \_\_ *as well as* \_\_, *as far as* \_\_, *for* \_\_, the importance of \_\_, and with the development of \_\_, appear far more frequently in CLEPSC than in the reference corpus, but typical sequences in native data, such as *to \_\_ this up*, *as I \_\_*, and *for better or*

*(for) worse*, are rarely used in CLEPSC. Much research needs to be conducted as to the use of these chunks by Chinese college students. Nevertheless, in the following discussion, I'll just focus on the chunk *as far as* \_\_.

Although this chunk is quite common in CLEPSC and has an occurrence of 20 times, it occurs only 7 times in the reference corpus. In CLEPSC, this chunk only covers three variations including *as far as I am concerned* (13 times), *as far as I know* (4 times), *as far as I can see* (3 times). In the reference corpus, however, the number of the variations of this chunk is seven, the same number its frequency in the reference corpus. The variations include *as far as we can see*, *as far as law practice is concerned*, *as far as this is concerned*, *as far as that subject is concerned*, etc. In CLEPSC, this chunk nearly invariably appears in clause-initial positions to begin an utterance of opinions. To varied degrees, the chunk has become a tool for students to introduce their views and judgments. As illustrated in the samples from CLEPSC below,

speakers' opinions are almost always conveyed or expressed in the wordings on the right of the chunk.

(3) <sp1> As far as I am concerned, this idea is radically wrong. <sp1/>

<sp2> As far as I am concerned, we can work out better solutions. <sp2/>

<sp3> As far as I am concerned, only a combination of both can help us conquer future challenges. <sp3/>

The abovementioned samples show that Chinese college students use this chunk merely to introduce their opinions. However, native speakers can use it to realize a myriad of pragmatic functions whereas barely deploy it to introduce their own perspectives but more to make statements of facts, as shown in the samples below:

(4) <sp1> I wanted the position of the man I think is to be the most envied, as far as the ability to do good is concerned. <sp1/>

<sp2> As far as that subject is concerned, the danger is great to America. <sp2/>

<sp3> As far as they're concerned, no one handed them anything; they built it from scratch. <sp3/>

The reason why native speakers rarely use this chunk to introduce opinions while Chinese college students do can be analyzed using Economy Principle. It refers to achieving the maximum effect with the least effort, i.e., conveying more information with the least possible time and energy. It is observed in both static and dynamic aspects of languages. In public speeches, Economy Principle is extremely crucial, insofar as the speaker intends to convey more information in a relatively short period, while the audience's ability to identify and memorize those points in the speech is very limited. Therefore, succinct language is preferred by almost every great speaker to achieve identification with the audience (Burke, 1969).

In this study, Economy Principle can be simply put as "we should delete words which perform no solid function in the sentence or add no meaning to the speech." Native speakers observe this rule consciously or unconsciously in their speeches. Long chunks like *as far as I am concerned*, *as far as I know* are almost always replaced with shorter chunks expressing similar meanings, such as *in my view*, *in my book*, etc. On

the other hand, Chinese college students seem to show little concern for this principle. They are more dependent upon chunks like *as far as I am concerned* for the following three possible reasons: (1) In Chinese college students' view, phrases like *as far as I am concerned* is considered advanced and idiomatic; substituting them for the ones like *in my opinion* is wrongly perceived as the demonstration of high English proficiency. (2) This kind of phraseology can allow for more time for the students to organize subsequent ideas, and thus increase speech fluency. (3) Speakers are usually asked to speak for two to three minutes in the impromptu speech section of an English public speaking contest. Speeches either too long or too short will lead to the deduction of scores. A good many of speakers have difficulty in delivering an impromptu speech which is long enough. Thus, using longer chunks might be a way for them to meet the time requirement.

## 5. Conclusion

In this paper, I have investigated the phraseological features of Chinese college student oral English in public speaking. The major findings of this study are:

Firstly, in terms of the overall distributions, the use of phraseology by Chinese college students bears a resemblance to that by native speakers across the four structural types of chunks, i.e., polywords, institutionalized expressions, phrasal constraints, sentence builders. It has been found that Chinese college students are not able to proficiently use these two types of chunks, which play a significant part in reducing the stress of immediate analysis and language decoding as well as enhancing the language naturalness. Their failure to pick these pragmatically specialized chunks might mar the pragmatic quality of discourse and lead to dysfluency in speech.

Secondly, in terms of functional categories of the chunks, there are both similarities and differences across the two groups. Similarities include, among others, the following aspects: (1) There are a host of different paradigmatic slot/filler variations for sentence builders, though they usually have simple syntagmatic frames. (2) Certain chunks are employed more often to achieve specific speech act functions under specific contexts or circumstances. (3) Fluency devices are used to allow for more time of thinking. On the other hand, some distinct



differences have been found across the two groups: (1) Many chunks used by Chinese college students, such as sentence builders and polywords, are large in number but small in type compared with native speakers. (2) Chinese college students tend to be more dependent upon several chunks, such as *my topic (today) is (that) X, good morning/afternoon, as far as I am concerned*, etc., which indicates not only a lack of diversity, but also Chinese college students' misconception of what chunks are idiomatic and appropriate in specific contexts. (3) Some socio-pragmatic differences have been observed. For example, Chinese college students prefer to use evaluators such as *it's a good idea to X*, whereas native speakers like to use *I'm a great believer in X*, with the word "believer" bearing religious implications.

The findings of this study can provide significant pedagogical implications for teaching spoken English, especially English public speaking. The most striking contrast across the two groups is that Chinese college students employ a host of chunks that rarely occur in the native data. The underlying reason, perhaps, is that Chinese college students perceive public speaking as reports to the audience, rather than communication with them. It is essential to understand how the different parts of language (chunks) function as parts of a discourse in learning how to make a speech (Nattinger & DeCarrico, 1992). The practical classroom teaching can be conducted in the following way: In the first place, a teacher can tell his students to consider delivering speeches as communicating with the audience. At the outset, the teacher can allow students to use whatever chunks they like. Later, the students need to be told that some chunks can be substituted with more appropriate ones. For example, in teaching topic markers, one would find the lexical phrase, *the point I'm trying to make is (that) X*, more natural than *my topic (today) is (that) X*. The next step would be to help the students gain fluency of these more idiomatic chunks. Pattern practice drills could help students gain fluency with some fundamental fixed routines first (Peters, 1983). The teacher's goal would be to employ such drills to build confidence and fluency while not overdoing them to the point of incurring mindless exercise. Subsequently, the teacher can use simple substitution drills to introduce students to variation of these fundamental phrases, demonstrating that the chunks learned

before were not fixed sequences, but rather patterns with open slots. In this regard, chunks with several slots, like sentence builders, are more ideal types for teaching. As students are more competent at employing the phrases, the range of variation can be expanded, allowing them to gain a more profound insight into the essence of those chunks. The objective would be for students not only to remember the chunks provided in the lectures, but to learn to segment and create new patterns on their own. Then students can put chunks into good use in English public speaking.

It must be admitted that this study is not free from limitations. Firstly, it should be noted that a few chunks might be left out during the process of screening, shrinking the number of chunks investigated in this study. Secondly, as the selected speeches are delivered by students who are qualified to compete in speech contests, i.e., those with relatively high English proficiency, the size of the corpus is therefore limited. Thus, the data may not accurately reflect the hallmark of the Chinese college student oral English in English public speaking. Further work needs to be done to establish whether similar formal and functional features of the use of phraseology can be found among Chinese college students with lower English proficiency.

**Funding:** None.

**Declarations of Interest:** None.

## References

- Ädel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for specific purposes*, 31(2), 81-92.
- 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.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at...: Lexical bundles in university teaching and textbooks. *Applied linguistics*, 25(3), 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., Finegan, E., & Quirk, R. (1999). *Longman grammar of spoken and written English* (Vol. 2). London: Longman.



- Boers, F., & Lindstromberg, S. (2005). Finding ways to make phrase-learning feasible: The mnemonic effect of alliteration. *System*, 33(2), 225-238.
- Burke, K. (1969). *A Rhetoric of Motives*. Berkeley: University of California Press.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language learning & technology*, 14(2), 30-49.
- Cortes, V. (2002). Lexical bundles in Freshman composition. *Using corpora to explore linguistic variation*, 9, 131-145.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for specific purposes*, 23(4), 397-423.
- Ding, Y. R. & Qi, Y. (2005). Use of Formulaic Language as a Predictor of L2 Oral and Written Performance. *Journal of PLA University of Foreign Languages*, 28(3), 49-53.
- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual review of applied linguistics*, 32, 17-44.
- Firth, J. R. (1951). Modes of meaning. *Essays and Studies (The English Association)*, 118-149.
- Hyland, K. (2008a). Academic clusters: Text patterning in published and postgraduate writing. *International Journal of Applied Linguistics*, 18(1), 41-62.
- Hyland, K. (2008b). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4-21.
- Jespersen, O. (1924). *The philosophy of grammar*. London: Allen & Unwin.
- Leech, G. (1983). *Principles of Pragmatics*. London: Longman.
- Nattinger, J. & DeCarrico, J. (1992). *Lexical phrases and language teaching*. New York: Oxford University Press.
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and communication*, 191, 225.
- Peters, A. M. (1983). *The units of language acquisition* (Vol. 1). CUP Archive.
- Römer, U., & Arbor, A. (2009). English in academia: Does nativeness matter. *Anglistik: International Journal of English Studies*, 20(2), 89-100.
- Simon, H. A. (1974). How Big Is a Chunk? By combining data from several experiments, a basic human memory unit can be identified and measured. *Science*, 183(4124), 482-488.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16(2), 180-205.
- Wood, D. (2002). Formulaic Language in Acquisition and Production: Implications for Teaching. *TESL Canada Journal*, 20(1), 1-15.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.