UNIVERSITY OF Southampton

University of Southampton Research Repository
ePrints Soton

Copyright © and Moral Rights for this thesis are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

http://eprints.soton.ac.uk

UNIVERSITY OF SOUTHAMPTON

FACULTY OF HUMANITIES

MODERN LANGUAGES

AN EXPLORATION OF FORMULAIC LANGUAGE IN CHINESE UNIVERSITY STUDENTS' WRITTEN TEXTS

by

Jiaoyue Chen

Thesis for the degree of Doctor of Philosophy

September 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF HUMANITIES

Modern Languages

Thesis for the degree of Doctor of Philosophy

AN EXPLORATION OF FORMULAIC LANGUAGE IN CHINESE UNIVERSITY STUDENTS' WRITTEN TEXTS

Jiaoyue Chen

Over the past few decades, there has been an increased interest in the formulaic aspects of languages, including English. There has also been work conducted into the learning and teaching of formulaic language and its use by non-native speakers. Despite the increase in English language teaching in China, there has, however, so far not been any combined research into the learning, use and teaching of formulaic language in the Chinese EFL context. This study addresses this gap by investigating the written texts of Chinese university students and the learning, use and teaching of formulaic language in this research context.

As background to this study, an overview of existing research on formulaic language is firstly introduced, and then the rationale for this study, investigating formulaic language through student written texts, is established by positioning the role of written language in second language research and the relation between formulaic language and genre analysis studies. After these, specific background information on the EFL context in China is presented by supporting the claim that the EFL students in the study are seen as language learners as well as writers and users of English. The research questions that this study sets out to answer are the following:

(1) To what extent do Chinese university students use formulaic language in their written English?

a) What are the main structures of formulaic language used by these learners?

b) What are the main discourse functions of formulaic language used by these learners?

c) What is the relationship between the distribution of structural and functional categories of formulaic language in the learners' written texts?

d) How is the formulaic language used differently in the written texts of Year1 and Year 3 university students?

(2) What do Chinese university students perceive formulaic language to be?

a) To what extent is this perception different in Year 1 and Year 3 students?

(3) How do Chinese university students perceive the learning, use and teaching of formulaic language?

The results of this research will present formulaic language use in student written texts, and link this use to students' self-reported processes, strategies and sources of formulaic language learning and use. Nevertheless, student reflections on the teaching of formulaic language in the research context will be introduced in order to address the research questions thoroughly.

A mixed methods research design is employed in this study. The fieldwork took place during one semester (16 weeks) at a Chinese university. The participants were 83 students from the Year 1 and 73 students from Year 3 groups in the foreign language department. The main sources of data were firstly, students' written texts and secondly, in-depth interviews with 12 informant writers.

The findings of the research present a well-rounded description of formulaic language use in Chinese university students' English written texts, by analysing and comparing the distribution of structural and discourse functional categories in the formulaic strings identified by the student perceptions and through corpus linguistic methods. Also, triangulation of the textual data collected from the written texts and perceptual data gathered from interviews shows some discrepancies regarding the perception of

formulaic language in English among students and in researchers in the field. This thesis ends with a discussion of the implications and limitation of the present study, and directions for future research on formulaic language in the EFL context.

Table of Contents

List of Figures

List of Tables

DECLARATION OF AUTHORSHIP

I, JIAOYUE CHEN

declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

AN EXPLORATION OF FORMULAIC LANGUAGE IN CHINESE UNIVERSITY STUDENTS' WRITTEN TEXTS

.....

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. [Delete as appropriate] None of this work has been published before submission [or] Parts of this work have been published as: [please list references below]:

Signed:…… ……..

Date:…… ……..

ACKNOWLEDGEMENTS

Completing this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people.

First and foremost, I want to thank my supervisor Dr Julia Hüttner. It has been a great honour to be your student. I appreciate all your guidance and support through my PhD. I am thankful that you taught me to be independent, diligent, patient, open-minded, engaging and to stick to the commitment. I am also thankful that you are the excellent example yourself have provided as a successful woman researcher and practitioner in Applied Linguistic and Language Education. I would also like to thank Dr Alasdair Archibald, you have inspired my interest in the written language research and supervised my MA dissertation on formulaic language in 2011. I want to thank you for being supportive and understanding during my study in the UK. I would also like to thank Prof Rosamond Mitchell who gives me advice and continues to resonate at every stage of this PhD.

I sincerely appreciate advice and comments offered by the examining committee: Dr Lesley Gourlay and Prof Vicky Wright. I am sure your invaluable feedback improves the final version of my thesis.

I am also grateful for the academics and fellow researchers in the field of formulaic language and second language research. My special thanks goes to Professor Alison Wray for bringing me into the FlaRN (Formulaic Language Research Network) and giving me continuous advices; and to Professor Florence Myles for your insightful comments and encouragement.

Meanwhile, I would like to thank the staff members in the Modern Languages department. I have not only acquired knowledge and skills in Applied Linguistics and related areas, but also have learned the qualities of being a good researcher and an inspiring teacher from the staff team, my thanks specially go to Jenny, Lu, Will, Glyn, Laura, and Karin. I would also express my appreciation to Dr Louis Rayment, who proofreads the whole thesis.

I should truly thank all the staff members in the foreign language department at the Hubei University of Automotive Technology who made my fieldwork in China possible and offered every possible assistance during my time there. I must also add further thanks for Professor Xianyu Yang, Zhen Li, Juan Xu and Qiong Chen, for their help with the collection of students' written texts. At the same time, I owe a debt of gratitude to the students from T1093 and T1293 who agreed to take part in the research. Without you, this thesis would not be the same.

I am also grateful to all my English language teachers, Xinrong Pan (my auntie, also my first English language teacher), Hongping Cheng, Yueqin Hu, Guifang Li, Shiyu Wang. I want to thank you all for opening the window of English world for me and making me inquisitive in English language. Thank you very much for your teaching and continuous support!

I have to say thank you to all my family and friends both in the UK and in China, very nice to know you all and even better to have you in my life all these years!

Last but not least, I owe the deepest gratitude to my father and mother. Without you, everything I have achieved is impossible. Thanks for your unconditional love, endless support and unlimited patience and everything you both have done for me. I would like to dedicate this PhD thesis to my parents. 谁言寸草心, 报得三春晖。

ABBREVIATIONS

Adj. P: Adjective Phrase

Adv. P: Adverbial Phrase

CET 4: College English Test Band 4

CET 6: College English Test Band 6

CLEC: The Chinese Learner English Corpora

EAP: English for academic purposes

EFL: English as a foreign language

ESL: English as a second language

ESP: English for specific purposes

FEI: Fixed expressions and idioms

ICE: International Corpus of English

ICLE: International Corpus of Learner English

IELTS: International English language test system

L1: First language

L2: Second language

MI: Mutual information

MWU: Multi-word unit

NP: Noun Phrase

PP: Prepositional Phrase

SFG: Systemic Functional Grammar

TEM-4: Test for English Major Band 4

TEM-8: Test for English Major Band 8

VP: Verb Phrase

WECCL: The Written English Corpus of Chinese Learners

Chapter 1 Introduction

1.1Background

My initial motivation for studying formulaic language was very personal. Several years ago, when I took training courses for the IELTS, my writing tutor told the class not to start writing by using a phrase like "with the development of" in tests, since this kind of terminology was used too frequently and the examiner might recognise that the author was Chinese, or at least, with limited proficiency level of English. He continued to introduce some other phrases that should be avoided, such as *for example*, *different from*, and *I think*. When I attended the English writing course at university, as an English major student, the teacher also gave us such suggestions. The reason given was always the same: these items were just used too frequently by students. From then on, the questions "how can people recognise me as a non-native speaker of English just by those phrases used in one piece of my writing?" and "why do we have to learn these phrases if they may hinder English language performance?" have intrigued me.

Dechert (1983:156) uses the phrase "islands of reliability" or "fixed anchorage points" to refer to this language use, and explains that learners' repertoires for introducing arguments and points of view are very restricted, therefore they cling to certain fixed phrases and expressions which they feel confident using. Hasselgren (1994) also refers to these particular chunks functions as "lexical teddy bear". But what exactly are these fixed phrases and expressions?

The initial inquiry starts with looking for a proper term for this language phenomenon. Indeed, there is a number of terms used to refer to formulaic features in languages, such as idioms, collocations, fixed expressions, lexical bundles, lexical phrases, phraseology, and multi-word combinations. Wray (2002) proposes the umbrella term "formulaic sequence" to refer to these continuous or discontinuous prefabricated chunks retrieved from memory as a whole, without having been generated or analysed by the language

grammar. In this study, formulaic language is used as an inclusive term to refer to formulaiciy in language use.

Erman and Warren (2000), calculated that formulaic language of various types constituted 58.6% of the spoken English discourse that they analysed and 52.3% of the written discourse. Using different criteria and procedures, Foster (2001) found that 32.3% of the unplanned native speech that they analysed was made up of formulaic language. Nattinger and DeCarrico (1992), also argue that formulaic language is ubiquitous in language use. Based on such previous studies, despite certain discrepancies among the research results on the proportion of the English language taken up by formulaic language, it can be seen that formulaic language is widely used in English.

Formulaic language has long been recognised as having an important role in language acquisition in relation to processing, fluency, and idiomaticity. Therefore, it has also been the focus of a wide range of studies across psycholinguistics (N. Ellis, 2012; Robinson and Ellis, 2008) and corpus linguistics research (Biber, et al., 2004; Hyland, 2004; 2008a, b; Simpson-Vlach and Ellis, 2010).

Corpus linguistics has demonstrated that much of communication makes use of formulaic language, therefore language is rich in collocational and colligational restrictions, and the phrase is the basic level of language representation where form and meaning meet with great reliability (O'Donnell, et al., 2015). There are two extraction methods used in corpus linguistics studies: co-occurrences and recurrences. The occurrence is investigated by the mutual information (MI) statistic score (Lorenz, 1999), which highlights co-occurrences that are not especially frequent, but are closely associated and are thus likely to be very salient for native speakers (Simpson-Vlach and Ellis, 2010). However, Paquot and Granger (2012) note that this extraction method, which is based on statistically significant co-occurrences, has seldom been used in the learner corpora research, probably because the associated measurement tends to become unreliable with low-frequency data (Evert, 2008), and the learner corpora have

until recently been of very limited size. The other extraction method used to identify recurrent patterns by retrieving all repeated multi-word combinations is called lexical bundles research (Biber et al., 1999; 2006; Biber and Barbieri, 2007; Hyland, 2008a, b). These studies usually provide a structural and functional description of lexical bundles, and describe English academic writing in terms of grammatical compression, syntactic elaboration and degree of explicitness (Biber, 2006; Biber et al., 1999). Biber and Barbieri (2007: 265), suggest that each academic genre displays "a distinct set of lexical bundles, associated with [its] typical communicative purposes". This also supports the claim that every genre of EAP (English for Academic Purposes) and ESP (English for Specific Purposes) has its own formulaic language, and learning to be effective in the genre involves learning the relevant formulas (Swales, 1990).

Formulaic language has also been studied using a psycholinguistics approach. Psycholinguistic research independently investigates the psychological reality of formulaic language using experimental methods (O' Donnell et al., 2015). With this approach, it is possible to demonstrate language users' sensitivity to the frequencies of occurrence of a wide range of different linguistic constructions, and therefore to provide a clear statement of the influence of each usage event, and the processing of its component constructions on the learners' system (N. Ellis, 2008). Research in this strand has also offered evidence that formulaic language is learned "as wholes" and not as strings of individual words, that formulaic language is learned incrementally and that fluent language users inhabit have an ample repertoire of memorised language sequences (N. Ellis, 2008; Li and Schmitt, 2009).

There has been heated discussion about the compatibility between the speaker-internal approach (psycholinguistic account) and the speaker-external approach (corpus linguistics research on frequency account) to formulaicity in language production. As Flowerdew (2012) argues corpus linguists tend to use expressions containing the string "phrase", whereas psycholinguists use expressions with the string "formula". This is later explained in Myles and Cordier's (2016) study, and they clarify that the speaker-external

approach refers to studies on formulaicity in a language outside of the speaker, thus these studies usually focus on the formal aspects, frequency of occurrence in corpora or pragmatic functions; whereas the speaker-internal approach refers to the research on psycholinguistic salient units for a speaker, either those units are stored holistically or retrieved with greater efficiency than other linguistic strings. Nevertheless, as Schmitt, et al. (2004) note, word strings come from corpus analysis (but which may or may not be stored holistically in the mind as whole units); whereas the formula may or may be not identifiable through the corpus. They also draw the conclusion that the two approaches are actually complementary (Schmitt and Carter, 2004). However, there are few studies on formulaic language that are based on both theoretical research roots.

One recurrent issue in the existing discussions of formulaic language has been its connection to native-like language use. Researchers such as Foster (2001), use native speaker norms to judge non-native speakers' formulaic language use. For example, a comparison has been made between native and non-native speakers' corpora in terms of formulaic language frequency, and the conclusion has been drawn that non-native speakers over- or under-use formulaic language and fail to understand the pragmatic functions according to English conventions (Chen and Baker, 2010; Granger, 1998). However, the notion of the native has also been problematic and has been challenged by a number of researchers (e.g. Jenkins, 2000; 2007; Kirkpatrick, 2007; Seidlhofer, 2006).

First, there is no research that could provide a workable and rational distinction between a native and non-native English speaker, as there is no exact definition of a native speaker to which everyone subscribes based on a substantial body of literature (Galloway and Rose, 2015: 202). On the contrary, many researchers claim that the term "native speaker" is very misleading. For example, Paikeday (1985:25) notes that the native speaker "exists only as a figment of a linguist's imagination" and he preferred the term "proficient user". Similarly, Rampton (1990), suggests the replacement of terms "native" and "non-native speaker" with expert and novice respectively, with the

intention of placing emphasis on language expertise rather than nativeness in the conceptualisation of language competence. In addition, Modiano (1999), argues that the primary criterion for classifying different speakers of English should be language competence, rather than nativeness, as language competence should not be something owned solely by native speakers.

Second, it has been considered problematic to pass native-speaker judgements on appropriate usage. For example, Jenkins (2000: 160) claims that "there really is no justification for doggedly persisting in referring to an item as "an error" if the vast majority of the world's L2 English speakers produce and understand it". In other words, traditional native and non-native distinctions in the English as a foreign language (EFL) context place too much emphasis on accounting for language forms rather than considering all the contextual factors in the communicative event and language users in their own right.

From the language learning perspective, previous research on formulaic language provides convincing evidence that learners unusually begin with formulaic language ties in real life interaction and proceed to "break down" or analyse language and eventually use it creatively (Yu, 2015). It also shows that it indeed is a great support in the early stages of L2 learning for most types of learners (i.e. Myles, et al., 1998; Weinert, 1995). Nevertheless, Warren (2005:38), claims that native-like mastery of idiomaticity is difficult to foreign language learners. As for classroom-taught teenagers and adults, especially, when they have achieved a reasonable command of the L2 lexicon and grammar, their command of formulaic language appears to be lagging behind (Wray, 2002). However, there is little research focusing on classroom learners (at the same time, language users) and analysing the reasons why their command of formulaic language lags behind. As Wray (2002) notes that research offers little insights to why such an approach may appeal or what the spontaneous appearance of formulaic language in interlanguage might specify or reveal. On the other hand, previous research with varied corpora has demonstrated that formulaic language is associated with

language proficiency level and expertise in academic writing, but the relationship remains unclear (Biber, et al, 2004; Chen and Baker, 2010; Cortes, 2004; 2008; Huang, 2014, 2015; Hyland, 2008a, 2008b; Leedham, 2011; Li and Schmitt, 2009).

So far, formulaic language, as a ubiquitous language phenomenon has been briefly introduced by reviewing the main research approaches: the psycholinguistics approach and the corpus linguistics approach. As the primary aims of this research are to explore and describe formulaic language in Chinese university students' English written texts. In order to produce a well-rounded description of formulaic language use, both approaches to formulaic language will be considered in this research. In other words, the frequency-based recurrences and advantaged processing units will be taken as the independent premise in the working notion of formulaic language for this study, regardless of English native speaker norms. I then propose the definition of formulaic language in psycholinguistic and corpus linguistics constructs respectively as follows:

> 1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner, and it is not constructed by words when learners use it, regardless of language native norms during identification:
> 2) Linguistic cluster, which meets a series of selective criteria in a corpus;

In the next section, the research purposes will be further clarified and research questions will be presented to guide this study.

1.2 Research purposes and questions

The primary aim of this study is to provide insightful information about the learning, use and teaching of formulaic language by Chinese university students through their English written texts. Drawing on the related literature on formulaic language, this study first attempts to offer a description of Chinese EFL learners' formulaic language use in terms of linguistic structures and discourse functions, by answering to the first research question and its set of sub-questions (RQ 1a to 1d). Following this, I will then explore in

greater detail how the students conceptualise formulaic language with research question 2 and further examine the differences across groups (RQ 2a). Finally, attention is paid to the investigation of student perspectives on the learning, use and teaching of formulaic language at the research site. To summarise, the research questions posed in the study are:

(1) To what extent do Chinese university students use formulaic language in their written English?

a) What are the main structures of formulaic language used by these learners?

b) What are the main discourse functions of formulaic language used by these learners?

c) What is the relationship between the distribution of structural and functional categories of formulaic language in the learners' written texts?

d) How is formulaic language used differently in the written texts of Year 1 and Year 3 university students?

(2) What do Chinese university students perceive formulaic language to be?

a) To what extent is this perception different in Year 1 and Year 3 students?

(3) How do Chinese university students perceive the learning and use of formulaic language?

Through addressing these research questions, this research will contribute to a better understanding of the use, learning and teaching of formulaic language in the EFL context, by drawing on a joint view of psycholinguistic and corpus linguistics. Also, it will identify directions for further research in this field.

1.3 Structure of the thesis

This thesis consists of 8 chapters. Chapters 2 to 4 mainly provide a review of the relevant literature concerning three areas pertinent to this study: formulaic language, the learning and use of formulaic language in second language learning, and current EFL teaching and learning in China. More precisely, chapter 2 reviews the conceptualisation of formulaic language from different research positions. It provides a detailed examination of the terminologies used across formulaic language research, and two main approaches to the identification of formulaic language: the corpus linguistics approach and the psycholinguistics approach. This is followed by a review which outlines the importance of formulaic language in related areas, including grammar, language processing, and communication, and specifies the contributions of previous research.

I mainly discuss the literature in relation to my own theoretical framework in Chapter 2, which draws on both corpus linguistic and psycholinguistics approaches to formulaic language. I argue here that while formulaic language is characterised by a certain level of conventionalisation, it does vary between individuals and importantly across speaker groups, including non-native speakers of English. Therefore, it is not justifiable to solely use English native speakers' intuitions to judge the formulaic language used by non-native speakers. In my study, thus, I used both speaker-internal and speaker-external approaches to constructing the definition of formulaic language, which draws on both psycholinguistic and corpus linguistic approaches:

> 1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner, and/or retrieved with greater efficiency than other linguistic strings at the time of use, regardless of native language norms during identification;
> 2) Linguistic cluster, which meets a series of selective criteria in a corpus, usually including the length of the clusters, threshold frequency, and text dispersion.

Chapter 3 establishes the rationale for researching the learning and use of formulaic language through EFL student written texts. This chapter starts by reviewing the essential role of formulaic language in second / foreign language learning, including the roles of formulaic language in second /foreign language learning, formulaic language learning, and second language learner strategies. It then examines previous research on formulaic language use in student written texts, providing the reasons why written language has been chosen as a medium for studying formulaic language in this research, followed by a review of previous studies on formulaic language in student written texts.

In this chapter, the researcher clarifies the shift in previous writing research from focusing on the surface linguistic feature to providing further explanations of language use in context. I argue that formulaic language cannot only be studied through learner written texts based on a traditional corpus-based approach, but must also take in ideas from genre analysis studies in order to obtain a rich description of student language use, especially the idea that formulaic language is conventionalised language use. By means of this, my analytic framework for studying formulaic language is further rationalised and presented.

Chapters 4 to 8 present the original empirical study of this thesis. Chapter 4 provides situational information on the student writers, as well as the research site. At the beginning of this chapter, I draw on Kachru's (1986) circles of English language users and make clear the EFL position for the writers in this research. Then, the general picture of English majors in Chinese higher education is presented by examining the curriculum and the research on actual language teaching and learning situation. Some issues and concerns raised by the researcher which are pertinent to the case of China and then, a further description of a local institution is provided.

Chapter 5 first restates and analyses the research questions, and then introduces the relevant research design models for this study. After a description and justification of the mixed methods research models introduced, the chapter moves on to an explanation of the selection of participants, the researchers' role, fieldwork, research

instruments, as well as an overview of databases. Next, the date analysis and operationalization for the written texts and interviews are presented and discussed step by step. The triangulation of research instruments will also be provided at the end of this chapter. Following this methodology, this study will provide rich textual and perceptual data to address the research questions regarding the use, conceptualisation, learning, and teaching of formulaic language in the research context.

Chapter 6 presents the data analysis results and general findings for each method and further clarifies the purposes of each research method and data collection. The main research methods used in the present study are the corpus analysis of students' English written texts and in-depth interviews with informant writers. The results and findings of each research method are presented in turn. At the same time, the validity and reliability of data analysis and results are clarified. At the end of this chapter, the triangulation of data analysis is also presented.

Chapter 7 synthesises the results from Chapter 6 according to the three sets of research questions. It begins with specifying the nature of each set of research questions and each type of data. In this way, the data and results are organised to answer the questions in turn. First, it presents and discusses the textual data and offers a well-rounded description of the formulaic language used by Chinese university students in their own English written texts, in terms of the grammatical structure and discourse function. The group differences are also presented and discussed regarding the grammatical structure and discourse function. Next, the perceptual data and results regarding conceptualisation, learning, use and teaching of formulaic language are presented in order to examine formulaic language in students' perceptions, which is somewhat divergent from the prior studies. Finally, the researcher discusses the limitation of the present study critically and extensively.

Chapter 8 offers the concluding remarks for the thesis. It first reviews the rationale for the research topic and revisits the gap in the research areas of formulaic language, in second language learning and/or foreign language learning settings. Then it goes

through the three sets of research questions and explains the methods used to answer each question. Next, it presents the findings in relation to the research focus and research questions. Finally, the thesis provides some recommendations and suggestions for future study as well as noting some of the implications for the teaching of formulaic language, especially in an EFL setting.

Chapter 2 Literature review of formulaic language research

This chapter will provide a thorough review of the related on formulaic language, and identify the gaps in the research area, especially with regard to the EFL context. This chapter first introduces terminology issues regarding formulaic language. In this study, I use the term “formulaic language”, as it is a neutral term for research with various theoretical roots. I then examine the identification of formulaic language based on the speaker-external (corpus linguistics) and speaker-internal (psycholinguistics) approaches. Formulaic language is thus repositioned for the language users, regardless of English native speaker norms. Therefore, it is inclusively established based on two constructs for this study, which is an innovative standpoint. In addition, literature in relation to the significance of formulaic language in language grammar, language processing, and communication is reviewed and discussed.

2.1 Terms used in formulaic language research

As noted in the editor’s introduction to the special issues of the *Annual Review of Applied Linguistics* on formulaic language, Polio (2012: vi) clarifies that formulaic language means

> what has been called formulaic sequences, multi-word expressions, lexical bundles, interactional routines, language chunks and so on… and that not all authors define formulaic language in the same way.

This corresponds to Wray’s earlier book (2002), where she lists various terms used to describe aspects of formulaicity in language, as illustrated below:

Amalgams – automatic – chunks – clichés –co-ordinated constructions – collocations-complex lexemes – composites – conventionalised forms – F[ixed] E (xpressions) including I[dioms] - fixed expressions – formulaic language – formulaic speech – formulas/formulae – fossilized forms – frozen metaphases – frozen phrases – gambits – gestalt – holistic – holophrases – idiomatic – idioms – irregular – lexical simplex – lexical(sized)phrases – lexical(sized) sentence stems – listemes-multi-word items/units – multi-word lexical phenomena – noncompositional – noncomputational – non-productive – nonpropositional – phrasemes – praxons – preassembled speech – precoded conventionalised routines – prefabricated routines and patterns – ready-made expressions – ready-made utterances – recurring utterances – rote – routine formulae – schemata – semipreconstuctured phrases that constitute single choices – sentence builders – set phrases-stable and familiar expressions with specialized subsenses - stereotyped phrases-stereotypes – stock utterances – synthetic – unanalyzed chunks of speech-unanalyzed multiword chunks-units.

Figure 2.1: Terms used to describe aspects of formulaicity (Wray, 2002: 9)

Wray remarks that several terms are used to describe the same phenomenon; however, it is also evident that some of the terms shared across different fields do not mean entirely the same thing in all instances (Wray 2002:8-9). Based on this consideration, she proposes an umbrella term "formulaic language" with the working definition as follows:

> a sequence, continuous or discontinuous, of words or other elements, which is or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.

Wray then further explains the notion of "formulaic" and "sequences" respectively. Thus, the word "formulaic" carries with it some association of unit, of custom, and habit; while "sequence" indicates that there is more than one discernible internal unit, of whatever kind (ibid.). On the notion of formulaic sequence, Wray (2002:10) further clarifies that "this term aims to be as inclusive as possible, covering any research field, and intending to make reference easier". Weinert (2010) suggests that the term

"formulaic sequences" is devised for Wray's particular, theory-sensitive definition, although many recent studies have employed the label "formulaic" as an umbrella term and referred to specific manifestations of the phenomenon with additional labels. Therefore, the term "formulaic language" is used as a general neutral term within various research strands, including corpus linguistics, psycholinguistics, and second language learning.

As for formulaic language research based on the corpus linguistic approach, with its specific research purpose and focus on certain pre-defined linguistic units, terms describing formulaicity have been used interchangeably across research traditions. Corpus studies may help us to understand "how different speakers can become co-ordinated and focused around the norms that we recognise as the idiomatic language use of a speech community" (Stubbs and Barth, 2003: 34).

Corpus linguistics research operationalises the notion of frequency as a defining feature of formulaic language and offers insights into how frequent specific units are within one corpus. This is based on the assumption that if a sequence is frequent in a corpus, this indicates that it is conventionalised by the speech community, at least to some extent.

Followed this tradition, formulaic expressions have been termed as "lexical bundles", "clusters" or "N-grams" (Biber, 1999; 2006; Hyland, 2008a, b), "phraseology" (Cowie, 1998), or "lexical phrases" (Nattinger and DeCarrico, 1992) and "multi-word units" (Moon, 1998). These terms all refer to those combinations of three or more words that are repeated without change a set number of times in a particular corpus, or in a specific community. For example, Moon's (1998) concept of a multi-word unit (MWU) is identified as conventionalised and institutionalised in a speech community, that is, everyone must recognise it as a unit that regularly reoccurs in language and does so with the same meaning and an MWU is institutionalised when it is used in a similar and consistent way by a speaker from the community. Sharing this rationale, there are a number of studies investigating recurrent phraseological units, for example, lexical

bundles (or clusters) is one of the important areas of research, which is based on a corpus linguistics approach across different genres. Examples of studies that analyse lexical bundles in various genres and registers include Allen (2010), Chen and Baker (2010), DeCock (2004), Biber (2006), Biber and Barbieri (2007), Huang (2014; 2015), Hyland (2008a; 2008b), Leedham (2011), Ruth (2013) and Salazar (2011) (section 2.2.1 and 6.1.2).

Within the research based on a psycholinguistic approach, terms like "prefabricated chunks" (Myles et al., 1998) and "formulaic sequences" (Wray, 2002; Schmitt and Carter, 2004) are often used interchangeably. Studies in this strand have mostly focused on the psycholinguistic status of formulaicity in language. Therefore, ready-made, memorised chunks are taken as the primary feature of formulaic language. For example, Myles et al. (1998) defined the term "chunks" as follows:

> a multimorphemic unit memorised and recalled as a whole, rather than generated from individual items based on linguistic rules.

This notion contains similar points to Wray's (2002) definition, as both definitions focus on the holistic store and processing advantages in the psycholinguistic status of formulaic language. Similarly, Schmitt and Carter (2004) emphasise that psycholinguists and language acquisition specialists focus on aspects which determine whether sequences are used by individual participants and whether these sequences are mostly formulaic and stored as a whole in the participant's mental lexicon.

Based on the review of related terms that are mostly used in previous research on formulaic language, it is necessary to be explicit in the choice of terminology and operationalised definition in any specific research. As Wray (2008:8) clarifies, formulaic language can be used as the "neutral mass (uncountable) noun", while the formula is used as the "neutral countable noun". As discussed above, in each research stance there are certain favoured terms due to the different focuses that studies put on the features of formulaic language. As Schmitt and Carter (2004:2) state

> Although [corpus] linguistic and psycholinguistic definitions developed for different purposes, any satisfying description of formulaic sequences probably needs to draw on both perspectives.

However, as for the differences and overlaps in formulaic language studied by these approaches, there are no definitive answers. Some researchers already argue that the findings from formulaic and frequent account based research may overlap when they likely to be processed holistically or more efficiently[1] (Myles and Cordier, 2016). As for the divergent aspects, it is assumed that some idiomatic expressions or clusters, which are commonly found in corpora are not necessarily stored holistically in all native speakers or L2 learners. Conversely, some speakers will store some sequences holistically which are not particularly frequent or shared by other speakers (Myles and Cordier, 2016; Schmitt and Carter, 2004; Wray, 2002, 2008).

In order to uncover these overlaps and divergences, and to further reduce the bias from each approach, I propose the neutral term "formulaic language" (Wray, 2008; Wood, 2010) for this study at the onset. It will be used as a general cover term, to refer to the study of formulaic language which includes both a corpus linguistics approach and psycholinguistic elements. Thus, formulaic language in this research includes the following two independent constructs: multiword psycholinguistic units and linguistic clusters. They are elaborated as follows:

> 1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner, and/or retrieved with greater efficiency than other linguistic strings at the time of use, regardless of native language norms during identification;

[1] Some researchers have carried out studies to distinguish whether a psycholinguistic formulaic language is holistically stored or with processing advantage. As the focus of my study is not on distinguishing between these two aspects, I use a general term "psycholinguistic salience" to refer to the psycholinguistic status that includes these two aspects. Thus, this issue is ruled out of the discussion in this thesis. More discussion can be found in Myles and Cordier (2016) and Siyanova-Chanturia (2015).

2) Linguistic cluster, which meets a series of selective criteria in a corpus.

Nevertheless, in the data analysis procedure, I will use the specific terms ---- formulaic strings and clusters to refer to the types of formulaic language which are identified by different approaches: student perceptions and corpus linguistics extraction respectively.

2.2 Identifying formulaic language

Wray (2012) used the metaphor "the blind men and an elephant" to describe the dilemma in the identification of formulaic language. Thus, all researchers in this area acknowledge the existence of the elephant (in this case, it refers to formulaic language), however, each research strand only focuses on its biased aspects in formulaic language (this is like blind men touch the tusk, trunk, or any other part of the elephant). No one provided a constant picture of the whole elephant yet. Wray (2002:19) notes that her definition of formulaic sequences focuses on the manner of storage, thus it is an internal and notional one rather than external or observable. She also makes it clear that the relationship between the definition and identification of formulaic language is circular and identification relies less on formal definitions than the definitions rely on identification (ibid.). This means that a comprehensive definition is actually less possible for the operationalization.

As proposed in the last section, formulaic language, in this study, is further defined based on two research strands respectively:

1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner, and/or retrieved with greater efficiency than other linguistic strings at the time of use, regardless of native language norms during identification;

2) Linguistic cluster, which meets a series of selective criteria in a corpus.

It is noted that the notion of formulaic language in this study draws on both psycholinguistics and corpus linguistic. On the one hand, with reference to corpus linguistics research, this study takes the frequency count as the primary feature of

formulaic language and measures it across the language production from one corpus; on the other hand, from the psycholinguistics perspective, formulaic language featured with internal storage manner and/or processing advantage is marked as a feature of similar importance to frequency count in the other operational definition.

According to Wray's (2002: 19) study, there are two basic ways in which formulaic language can be detected and collected. One way is to use an experiment, questionnaire or other empirical methods to target the production of formulaic language as data. The other justifiable way is to collect general or particular linguistic material and then extract strings in some principled way, based on some criteria. She then discusses the five main methods used to detect formulaicity in language, namely, (native speakers') intuition, frequency, structures, phonological forms, and some other methods used in specific data types (i.e. using code-switching in second language research).

Wray (2008) includes further ways to identify formulaic language by deepening and widening the specific approaches, including frequency, phonological indicators, form, idiosyncrasies, spelling, intuition, published lists and corpora, and mixed methods. She then concludes the two approaches to the identification of formulaic language: speaker-internal and speaker-external. This is later explained by Myles and Cordier (2016), who clarifies that the speaker-external approach refers to studies on formulaicity outside of the speaker, therefore, terms like linguistic clusters, bundles are often used in the studies in this strands; whereas the studies employed the speaker-internal approach to examine psycholinguistic units for a speaker, terms thus such as strings and sequences are mostly used.

This study employs both approaches to investigate formulaic language in the students' English written texts. By using the speaker-external approach, i.e. the linguistic clusters extracted based on the corpus linguistics method; and the speaker-internal approach, i.e. formulaic strings the student writers in their own writing, this research aims to provide an accurate description of formulaic language use by EFL learners with respect to the linguistic structures and discourse functions.

The following discussion will mainly be concerned with these two methods as they are employed in this research: the speaker-external identification by the corpus linguistics approach and the speaker-internal identification by taking psycholinguistic salient into account. For the speaker-external approach, studies on the structural and functional description of formulaic language will be reviewed; followed by the discussion on the psycholinguistics status, i.e., intuition to formulaicity, of formulaic language in the speaker-internal approach.

2.2.1 The speaker-external approach

In general, leading researchers in corpus linguistics, such as Biber, et al. (1999), Hunston (2002), Sinclair (1991; 2004), and Stubbs (1996, 2001), all view a corpus as a collection of authentic language, either written or spoken, which has been compiled for a particular purpose. Corpus linguistic research concerns itself with the description and explanation of the structure and use of language, and corpora have helped provide usage-based descriptive grammars of English (Biber, et al., 1999).

Part of this descriptive grammar has been a description of formulaic language. Flowerdew (2012:9) argues that the key defining feature of corpus linguistics is an analysis based on empirical data, and it tends to be associated with the phraseological approach to language, which is an approach to studying multi-word units of language. In corpus linguistics, computer searches are conducted to establish the location of multiword units within text quickly.

Two extraction methods have been mainly used within the corpus linguistic approach to formulaic language: co-occurrence and recurrence (Paquot and Granger, 2012). Co-occurrence describes the co-selection of usually two lexical items, which may be contiguous, and it refers to the preference that words have for certain co-occurring words (ibid.). In traditional phraseological studies, this was referred to as “mutual expectancy” (Firth, 1957), and defined as collocations or strict collocations (e.g. verb-object, adverb-adjective, or adjective-noun). For example, you can *perform a task*, or *do*

a task, but you cannot *make a task*. Also, a general trend was discovered and explained by Nesslhauf (2005), in that learners rely heavily on certain collocations (such as *have a look*, *do work, take care of*). He called this range of collocations the "collocational teddy bear" because learners feel confident using them. However, as Paquot and Granger (2013) noted that this extraction method is based on statistically significant co-occurrences and it has seldom been used in the learner corpora research, the reasons are due to the limited size for the most learner corpora, the associated measurement tends to become unreliable with lower frequency data in the database. Therefore, this type of extraction method is ruled out for this study.

The other type of formulaic language investigated in learner corpora is recurrences, which can be defined as the repetition of a contiguous string of words of a given length, such as bigrams or trigrams (Paquot and Granger, 2012). These strings are also referred to as "clusters", "chunks", or "bundles", regardless of the complete grammar units (see section 6.1.4). In a sense, these bundles are simply an extension of the collocations (Hyland 2012), which appear more frequently than expected by chance. Furthermore, they help to shape meanings in a specific context and contribute to the coherence of a text. Also, they are "conventionalised building blocks that are used as convenient routines in language production" (Altenberg, 1998).

Wray (2002: 25) also notes that this is done on the basis of frequency counts and considers frequency as a "salient, perhaps even a determining factor" in the identification of formulaic language, since "the more often a string is needed, the more likely it is to be stored in prefabricated form to save processing effort, and once it is so stored, the more likely it is to be the preferred choice when that message needs to be expressed" (ibid.). Tremblay, et al. (2011) examined the extent to which lexical bundles are stored and processed holistically by using eye-movement tracking methods. They found that lexical bundles and sentences containing lexical bundles were read faster by the subjects than the control sentence fragments.

It seems to answer “the unspoken assumption” that recurrent clusters identified by corpus analysis are also holistic formulaic sequences in the mind (Schmitt and Carter, 2004). Schmitt and Carter have argued that recurrent word clusters, for example, lexical bundles, which come from corpus analysis, may not be stored holistically in the mind and that formulaic sequences are stored in the mind as whole units but may or may not be identifiable through corpus analysis (ibid.).

In addition, both Wray (2008) and Flowerdew (2012) point out that when using computer searches, the researcher must decide what will count and what will not, and set up the search accordingly. For example, for multiword strings, decisions have to be made about how big the string should be, and how frequent an association has to be in order to count. However, such frequency thresholds can be arbitrary, particularly, in terms of the size of the corpus, the desired quantity of data, and the size of frequency counts which make it possible to quickly identify formulaic language through computer-based analysis.

2.2.1.1 The form of formulaic language

The perspectives that researchers have taken on the description of the linguistic structural aspect of formulaic language are varied. Previous studies on formulaic language by Nattinger and DeCarrrico (1992), Moon (1998) and Schmitt (2000) provide structural classifications for multi-word combinations.

Nattinger and DeCarrico’s (1992) study can be taken as a starting point for establishing the linguistic structural descriptive framework of formulaic language. As the most influential work in the field, it lays the foundation for later research. It proposes a structural classification and four formal criteria to characterize these phrases: 1) length and grammatical status; 2) canonical or non-canonical shape of the phrase; 3) whether the phrase is variable or fixed; 4) whether the phrase is continuous or discontinuous, that is, whether it consists of an unbroken sequence of words or whether it is interrupted by variable lexical fillers (ibid.). Based on these criteria, formulaic language

can be further divided into four categories, these are polywords, institutionalised expressions, phrasal constraints and sentence builders.

Nattinger and DeCarrico (1992: 37) further note that the classifications are based on structural criteria, but they also take some account of the function by listing the particular functions of individual examples. However, they do not exhaust the range of functions associated with these units. For example, in the category of institutionalised expression, besides the structural description, they also give the functional description as

> "institutional expressions are proverbs, aphorisms, formulas for social interaction, and all of those chunks that a speaker has found efficient to store as a unit. They are used for quotation, allusion, or (frequently) direct use" (Nattinger and DeCarrico, 1992: 39).

They then explain that some institutionalised expressions may either be general phrases used by almost everyone in the speech community, such as *how do you do* and *how are you*; or may be more idiosyncratic phrases that an individual has found been to be efficient and pleasing way of getting an idea across, for example, *have a nice day*. According to Nattinger and DeCarrico (1992: 40), institutionalised expressions are "mainly continuous, but at times they could be a discontinuous pair, for example, *once upon a time…and they lived happily ever after*, and frame chunks of entire texts".

Based on this categorisation, as Wray (2002: 49) comments, this classification system suffers from a lack of clear delimitations and an over-reliance on description only. As she suggests that only a small subset of formulaic sequences is entirely fixed, whereas those which are not, legitimately permit insertions (ibid.). Therefore, the fixedness criterion does not sit well with the existence of semi-fixed sequences. Usually, there is no clear-cut distinction between each type based on the pure description on the fixedness. For example, the differences between phrasal constraints and sentence builders, phrases,

like *the* ____*er, the* ___*er* [comparator] can be put the categories of phrasal constraints and sentence builders in Nattinger and DeCarrico's (1992) study.

Later, Moon (1998) suggests that three criteria come into play when defining the linguistic features of formulaic language (fixed expressions and idioms, FEI): institutionalisation, fixedness, and non-compositionality. Amongst these feature, she clarifies institutionalisation as "the process of by which a string or formulation becomes recognised and accepted as a lexical item of the language" and she also notes that "it is a necessary but not sufficient condition for a string to be classifiable as a fixed expression and idioms" (Moon, 1998:7). In this book, she especially highlights the term "idioms". She divides the notion of idioms into narrow and broad use: the narrow restricts idioms to a particular kind of unit, which is fixed and semantically opaque, and sometimes these are called pure idioms, for example, *kick the bucket*. Often, grammatically-ill-formed items such as *by and large*, or transparent metaphors such as *skate on thin ice* are excluded; whereas a broad idiom is a general term for many kinds of multi-word items, whether semantically opaque or not. She proposes that "dictionaries in the Anglo-American tradition often call fixed expressions 'idioms' making no further typological classification" (Moon 1998:4).

Schmitt (2000) has also provided description and classification for multi-word units and he has expanded the range of formulaic language to individual word level, by counting compound words as formulaic language. Thus, for Schmitt (2000: 99), compound words (for example, *blueberry*) are formulaic, as they are created when two or more words are combined to make a single lexeme and compounding is an important method of new word formation, which has led to the existence of large groups of compounds in English. In Conklin and Schmitt's (2012) study, they continue to argue that compounds are stored and processed as wholes, rather than generated and processed online. Likewise, results from previous research also indicate that individual words and compounds are both stored in the lexicon and that access to a compound can occur via individual words or the holistic representation (Pollastek, et al., 2000).

Recently, there has been a substantial increase number of studies drawing on the corpus linguistics approach to describing formulaic language, under various terms, such as "clusters", "multi-word strings", or "N-grams" (Biber, et al., 1999; Hyland, 2008a, b; Chen and Baker, 2010). Studies within this strand have followed Biber and his colleagues' typology on lexical bundles (Biber, et al., 1999), which mostly consist of phrasal combination with phrasal or clausal fragments. In line with this, Biber, Conrad, and Cortes (2004) classify lexical bundles that are used in the university register into four main types with sub-categories under each type. These four types of lexical bundles are lexical bundles that incorporate verb phrase (VP) fragments, lexical bundles that incorporate dependent clause fragments, and lexical bundles that incorporate noun phrase (NP)and prepositional phrase (PP) fragments (see section 6.1.2). A number of lexical clusters studies followed this structural classification (Bal, 2010; Leedham, 2011; Salazar, 2011).

The literature regarding the grammatical structure of formulaic language has been reviewed above, and it is hoped that this will provide some background information for the description of structural categories of formulaic language used by the students in their English written texts (see section 6.1.2). However, there are few studies examining formulaic language by taking both a corpus linguistics approach and a psycholinguistic approach into consideration. This leads to research question 1a, which aims to provide a well-rounded description of formulaic language using both approaches.

Nevertheless, Wray (2002) argues that there is no particular value in knowing that any function can be easily realised by any form, as there is no way of predicting. In other words, there is no sufficient or necessary link between certain forms and functions of formulaic language. Therefore, the relationship between form and function of formulaic language has been overlooked by previous research. This study sets off from two approaches to identification of formulaic language and tries to explore the relationship between the form and function of those identified formulaic items.

2.2.1.2 The functions of formulaic language

Focusing on function takes the position that formulaic language is an "expression whose occurrence is tied to more or less standardised communication situations" (Coulmas, 1981). As Nattinger and DeCarrico (1992: 59) note

> speaking a language means conversing in it, and comprehending a language means understanding phrases, not as isolated bits of grammatical structure, but as parts of the general ebb and flow of the surrounding discourse.

Some researchers have sorted out the functional categories of formulaic language in their own research (Cowie, 1988; Aijmer, 1996; Moon, 1992, Nattinger and DeCarrico, 1992).

Nattinger and DeCarrico (1992) offer functional categorisations for instances of formulaic language and classify them as social interactions, necessary topics, and discourse devices. Social interactional markers describe social relations and consists of categories of conversation maintenance and functional meaning relating to conversational purposes; necessary topics refer to the topics which learners are often asked or ones that are necessary for daily conversations. Discourse devices are lexical phrases that connect the meaning and structure of the discourse. They propose that these classifications are the most effective for pedagogical purposes, with a further aim to illustrate that formulaic language is effective and an essential signpost in sorting through materials that are often confusing and contradictory for second language learners (ibid.).

Cowie (1988: 134) identifies three function-based roles of formulaic language with reference to discourse, social and composites. In particular, Cowie notes that composites function as constituents in sentences (as objects, complements, adjuncts) and contribute to their referential, or propositional meaning. Similarly, Aijmer (1996) approaches functional aspects of formulaic language with a special interest in conversational routines and identifies three roles of formulaic language: 1) socio-

interactional functions, such as greetings, requesting, offering, or apologizing. 2) discourse roles. Aijmer explains two specific roles of formulaic language in discourse, one is to orientate the content, and the other is to categorise the text. 3) Attitudinal routines, are used to express the speaker's attitudes or emotions. Moon (1998) focuses on the functions of fixed expressions. Five functions of formulaic language are identified in her study: 1) informational, it is a vehicle for conveying new information and contributes to a discourse propositionally. Such as *in advance; by means of; on the grounds that*. 2) evaluative, it is used by speakers or writers to communicate their evaluations to their hearers or readers, for example, *a pain in the neck*; 3) situational expressions, i.e. greetings, apologizing; 4) modalising: conveying speaker's suggestions, advice and commands and it indicates modality, examples include *as we know it, at any price, at all*; 5) organisational functions used to organise the discourse and content.

Among four initially influential studies on the functional categorization of formulaic language: Aijmer (1996), Cowie (1988), Moon (1998), most functions discovered by the researchers share some common ground. In general, those functions concern three aspects: discourse organisation, interaction (between speaker and hearer) and evaluation (to show speaker's attitude). For example, the function of discourse organisation is mostly to structure and organise the discourse. Interaction indicates the relationship between the speaker and hearers. As for the evaluative function, this refers to the speaker's attitudes towards the topics.

Furthermore, Moon (1998: 217-218) suggests that the functions of FEIs can be related to Halliday's (1985, 1994) model of text components. The model views text in terms of its semantic stratification into ideational, interpersonal, and textual or textual components and it is a model for interpretation of ongoing dynamic discourse. She categorises the organizational and informational functions as subcategories under the ideational function; and puts situational, evaluative and moralizing under the interpersonal function. As for the textual component, she notes that it can be considered as "the

enable function", thus the FEIs are placed according to the topics and themes of the discourse (ibid.).

However, these studies do not focus on the registers of the discourse, mostly they examine the formulaic language in speech rather than written texts. There are no studies providing a well-rounded functional description of formulaic language in written texts before lexical bundles studies by Biber (2006) and Hyland (2008a).

Biber et al. (1999) coin the term lexical bundles to label "the most frequently occurring sequences of words" in a language or register. Lexical bundles are identified using special software on a large corpus of language. These expressions are groups of three or more words that frequently recur in a language or in a particular register. The selective criteria usually include the cut-off points for frequency (i.e. number of occurrences) and range (i.e. number of texts in which these bundles occur) (Biber and Barbieri, 2007; Biber, Conrad, and Cortes, 2004; Biber, Johansson, Leech, Conrad and Finegan, 1999).

The analytic framework for lexical bundles studies has been established based on Halliday's (1985; 1994) systemic functional grammar (SFG). SFG is a part of a social semiotic approach to language. Systemic refers to the view of language as a network of systems or interrelated sets of options for making meaning; functional refers to Halliday's view that language is as it is because of what it has evolved to do. Functions are used serve the structure and organisation of language at all levels and achieved via meta-functions. Halliday argues that this meta-functional organisation of language determines the form taken by grammar structure. In this sense, all languages have resources for constructing experience (the ideational component), resources for enacting human's diverse and complex social relations (the interpersonal component), and resources for enabling these two kinds of meanings to come together in coherent texts (the textual function) (ibid.). Based on Halliday's work, both Biber, et al. (1999), Biber (2006) and Hyland (2008a) proposed their own functional classification of formulaic language divided into three types of meta-functions (see section 6.1.2).

Biber (2006) outlines the three main functional categories for lexical bundles: stance expressions, discourse organisers, and referential expressions. Stance expressions relate to attitudes or assessments of certainty that frame some other proposition, examples include *I don't know what, the fact that the, I don't want to*. Discourse organisers reflect relationships between prior and coming discourse, such as *what I want to, those of you who, that's one of the*. Referential bundles make direct reference to physical or abstract entities, or to the textual context itself, either to identify the entity or single out some particular attribute of the entity as especially important. Examples of these are *the end of the, something like that*. Each of these categories has several subcategories associated with more specific functions and meanings. For example, stance bundles include epistemic stance bundles which comment on the knowledge status of the information (e.g. *I don't know if, the fact that the*) and attitudinal/modality stance bundles which express speaker attitude towards certain actions (e.g. *if you want to, you want me to*).

Studies within this framework have found that conversational and academic prose present distinctive distribution patterns of lexical bundles. For example, most bundles in conversation are clausal (for example, *can I have a*), whereas most bundles in academic prose are phrasal (for example, *the nature of the*) (Biber, Conrad, and Cortes,2004; Biber and Barbieri, 2007; Biber, 2006). Also, they have focused on comparisons between expert and non-expert writing (i.e. Biber, Conrad and Cortes, 2004). In general, Biber's classification emerged from a broader corpus of spoken and written registers which included casual conversation, textbooks, course packs, service encounters, and institutional texts.

Hyland further develops Biber's classification and establishes an analytic framework for lexical bundles used in written language across academic genres and disciplines (Biber, et al., 2004; Biber, 2006, Hyland, 2008a, 2008b). He identifies bundles within three broad foci of research topics, text, and participants. This classification is based on the view that writing is a social engagement, and also in an academic context reveals the

relationship between the writer and reader. Research-oriented bundles, which help writers to structure their activities and experiences of the real world, include subcategories of location (*at the same time*), procedure (*the use of the*), quantification (*the magnitude of the*), description (*the structure of the*) and topic bundles (*the currency board system*). Text-oriented bundles are concerned with the organisation of the text and its meaning as a message or argument, and include the following subcategories: transition (*in addition to the*), resultative (*as a result of*), structuring (*in the next section*) and framing signal bundles (*in the presence of*). Participant-oriented bundles, focus on the writer or reader of the text and include stance features (*it is possible that*) and engagement features (*as can be seen*) (Hyland, 2008a).

However, there are some other studies which originate from the function of formulaic language, in other words, the pre-established functions of formulaic language are used to identify formulaic language. For example, Durrant and Matthew-Aydınlı (2011) use a function-first approach for identifying formulaic language in academic writing, in which a corpus is first annotated for communicative functions and formulas are then identified as the recurrent patterns associated with each function.

In this section, literature on the external approach to formulaic language research has been reviewed, with regard to the formal and functional description. It provides theoretical standpoints for research questions 1a, 1b, and 1c. In the next section, previous research concerning psycholinguistic aspects of formulaic language will be reviewed. Some problems in the conceptualisation of formulaic language will be examined, in order to re-position formulaic language for this study.

2.2.2 The Speaker-internal approach

2.2.2.1 The role of intuition in identifying formulaic language

Intuition has been treated with suspicion in academic research. However, in the identification of formulaic language, although some other measures are primarily in use, intuition still tends to guide the design of experiments, the interpretation of results and

the choices of examples used in formulaic language studies (Wray, 2002:20). For example, idiomaticity can be defined straightforwardly in terms of members of the relevant speech community: an expression is idiomatic if it "sounds right", and is "regularly considered by a language community as being a unit" (Moon, 1998: 44). Or as Wray (2002) states, researchers, as members of their speech community, are often self-appointed arbiters of what is idiomatic or formulaic in their data.

According to Wray (2005), there are three levels of intuition: 1) ability to sense that something is ungrammatical, 2) ability to talk about grammar and 3) ability to make judgements about the data gathered. This seems to give a clear scope of intuition, however, as Wray (2002; 2008) argues there are some inherent problems in using intuition to identify formulaic language. First, in order to use intuition as a means of identification of formulaic language, only comparatively small data sets can be handled, which still necessitates laborious work. In contrast, frequency-based computer searches can easily handle corpora of any size. Secondly, an inherent inconsistency within the range of judgements made by an individual is inevitable, due to factors such as tiredness and unintended alternations in the judgement thresholds across time. Thirdly, there is probably no firm border in identified items, for example, if the researcher wants to analyse the phrasal structure, even if s/he employs identical criteria, the identified string could vary, especially in length. In addition, there are variations in judges, for example, different language proficiency levels and shared knowledge of formulaic language among the judges. As Sinclair (1991:4) suggests, human intuition about language is highly specific, and not at all a good guide to what actually happens when the same people use one language.

2.2.2.2 Criticisms of the role of intuition in formulaic language identification

Previous research has employed panels of native speaker judges to identify instances of formulaic language in learner language. For example, Foster (2001:83) formalised the procedures and tried to make it as reliable as possible, by using seven native speaker judges, who were all university teachers of Applied Linguistics with many years of

experiences in English as a foreign language. They were instructed to mark any language which they felt had not been constructed word by word but had been produced as a fixed chunk; or as part of sentence stem to which some morphological adjustments or lexical additions had been required. Additionally, these judges were asked to rely on their own judgement and not to consult anyone else. This study highlights the problematic aspects of shared knowledge of native speakers. Wray (2008:107) notes that intuition can come from different sources, as informants' intuitions can be used to find out what material they believe is formulaic for them. Wray also notes that if intuition is used as an approach to identify formulaic language, it is important for both researchers and judges to reach a shared sense of what constitutes a formulaic sequence (ibid). Like Foster's study, for members of the same speech community with shared knowledge, it might be possible to use intuition, as a measure of formulaicity.

For Sinclair (1991:4) the use of intuition is only useful for gaining insights into the nature of intuition itself, not the nature of language. Stubbs (2001: 71), on the contrary, suggests that in many areas of semantics and pragmatics, intuitions are strong and stable, across all native speakers, whether linguistically naïve or trained, and must be given the status of data. A compromise is offered by Flowerdew (2012: 44), who indicates that an alignment between a corpus and an individual's intuition would be likely when informants share the same set of usage norms. If the researcher and informant is a member of a particular discourse community, they could have an advantage as they have "the privilege of insider knowledge", thus they could use intuition as a reliable method to approach formulaic language.

However, from a cognitive perspective, even within a single linguistic community, native speakers are likely to have different sets of familiar formulaic language depending on their individual experience with language in various situations (Biber and Finegan, 1994; Biber, Conrad and Reppen, 1998), as the 'verbal repertoire consists of both dialectal and register variations' (Hymes, 1984). As Kaneyasu (2012) argues that an examination of formulaic language must take into account the linguistics variation, and in this sense, the

formulaic language analysis can be considered a kind of analysis which 'explores the link between linguistic expressions and social situation, with a view toward explanation' (Biber and Finegan, 1994:7).

Usually, it is clear that the researcher or the panel of native speakers of English who judge the formulaic language used by L2 or EFL learners, and it seems that as long as both researchers and judges reach a shared sense of what constitutes a formulaic sequence, the intuition-based approach could be reliable. Hereby, a more pressing problem looms in formulaic language research in second language acquisition or foreign language learning: is it justifiable to use native speakers' intuition to judge non-native speakers' use of formulaic language?

Wray (2002:143) points out that to know a language, one must know not only its individual words but also how they fit together. Part of this knowledge entails developing suitable rules to generate all possible grammatical utterances of the language, but another crucial aspect is coming to know which of the feasible grammatical utterances are idiomatic and nativelike. Prior to this, Willis (1990) proposes that a non-native speaker can only learn to use those which are the usual forms in a given speech community by observation and imitation. Most times, the term "idiomatic" is used interchangeably with "formulaicity" or "nativelike" (e.g. Wray, 2002; 2008), since idiomatic is used to describe the quality of language characteristics or turns of phrase that are commonly used by native speakers, as opposed to those grammatical sentences that a native speaker would not normally use (Pawley and Syder, 1983). Stubbs (1996:183) further comments that idiomatic language derives from a facility to use a much wider spectrum of formulaic sequences, one subset of which is idioms. A number of studies have defined and identified formulaic language based on the preference of native speakers, as shown in the following:

> "A combination of at least two words favoured by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalisation" (Erman and Warren, 2000:31).

"Native speakers' preferred combination of words" (Erman, 2009: 324).

It is noted that the notions of formulaic and idiomatic for a native speaker are seen as synonymous, whereas for a non-native speaker, they are not the same thing at all. The issue on whose formulaic language is usually lack of clarification or dominated by the native speakers' preference in the definition of formulaic language, both from psycholinguistic perspectives and corpus linguistics research strands.

2.2.2.3 Whose formulaic language is it?

Based on the arguments presented in the last section, I propose the question "of whose formulaic language are we investigating?" If a language item is said to be formulaic, then formulaic to whom? These questions are typically crucial when studying formulaic language in the ESL/ EFL context. First, formulaic language is produced by language users who are not native speakers of English, who normally do not have enough knowledge of idiomaticity in English language use. Second, the researcher, as suggested by Flowerdew (2012: 44) is usually playing two roles----that of a native speaker and that of the linguistic researcher. If the researcher does not share the same linguistic background with the language users, the shared knowledge of formulaic language between the researchers and the language learners might be very little. Especially in this research, formulaic language consists of two constructs:

1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner, and/or retrieved with greater efficiency than other linguistic strings at the time of use, regardless of native language norms during identification;

2) Linguistic cluster, which meets a series of selective criteria in a corpus.

The focus on only native-like formulaic sequences has been prevalent throughout the study of formulaic language. Pawley and Syder (1983) suggest that the formulaic language used by native speakers is not easy for learners to identify and master, and that its absence greatly contributes to learners not sounding idiomatic. Also, as

discussed earlier, native speakers have been frequently used as judges of what is or is not considered as formulaic (e.g. Foster 2001). As this study focuses on adult EFL learners, they study English in the classroom, we are dealing with a community of non-native speakers of English who share some characteristics in their language use. Wray (2002), for example, indicates that formulaic language has a positive role in the early stages of L2 acquisition for most kinds of learners, but highlights that by the time the learner has achieved a reasonable command of the L2 lexicon and grammar, their formulaic language appears to be lagging behind. Unlike the native speakers, learners generally do not know or have command over certain formulaic expressions unless they have actually encountered them before at a point in their learning when they had a chance to make sense of them. In 1989, Stevick provided some indication that successful language learners recognize the role of formulaic strategies. This includes deliberate mimicry of teachers' utterances, mechanical practice, and structured rehearsal of formulaic frames with different open slots, memorizing entire texts, and stock- piling sentences. However, as Wray (2002) notes, research offers little insight to why such an approach may appeal or what the spontaneous appearance of formulaic language in interlanguage might specify or reveal.

The present research aims to provide insightful information about the learning, use and teaching of formulaic language in an EFL context, as well as probing student conceptualisations of formulaic language. It is hoped that this study will offer a constant picture of formulaic language in an EFL learning context with in-depth insights, based on the notion that formulaic language can be formulaic for language users regardless of whether they are native speakers or non-native speakers of English. On the basis of these, the notion of formulaic language in this research is expressed through two constructs as follows:

1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner,

and/or retrieved with greater efficiency than other linguistic strings at the time of use, regardless of native language norms during identification;

2) Linguistic cluster, which meets a series of selective criteria in a corpus.

2.3 Roles of formulaic language in language use and learning

Formulaic language has been clearly recognised as an important and ubiquitous element of language learning and use. Formulae are used to facilitate language processing, help with fluency, and reduce the processing load in communication (Conklin and Schmitt, 2012). In this section, the importance of formulaic language will be revisited, to lay the foundation for further discussion about formulaic language in the research context. First, the theoretical significance of formulaic language will be reviewed, followed by a discussion of the role of formulaic language in language processing. Finally, this section reviews the value of formulaic language for communication.

2.3.1 Formulaic language and grammar

As for the relation between formulaic language and grammar, previous research has drawn on the theoretical significance by justifying formulaicity with reference to an external model of what language is and how it works. In the words-and-rules approach, a distinction is made between the lexicon (a collocation of memorised and stored forms) and grammar (a collection of rules that are applied to these forms). In general, it is accepted that we store representations of individual words in the mental lexicon. There is growing agreement that the lexicon also contains formulaic language, for example, *how are you* and *kick the bucket* (cf. Schmitt, 2000; Conklin and Schmitt, 2012). Conklin and Schmitt (2012) propose that the brain represents formulaic sequences in long-term memory, by-passing the need to compose them online through word selection and grammatical sequencing in a capacity-limited working memory (Pinker, 1999; Pinker and Ullman, 2002). Researchers in usage-based (Bybee and Hopper, 2001; Bybee, 2010; Goldberg, 2006; Tomasello, 2003) and construction-based models (Langacker, 1986) propose that the basic unit of language acquisition is a construction and that the task of

a language learner is to acquire a set of constructions that vary in size, complexity, and level of abstractness (Goldberg, 2006; Tomasello, 2003). These theories support the idea that all linguistic information is represented and processed in the same way, and that thus it should be similarly affected by frequency (Conklin and Schmitt, 2012). Therefore, language should be viewed not as a set of grammar rules, but as a statistical accumulation of experience that changes every time a particular utterance is encountered (N. Ellis, 2002).

The view that formulaic language is represented in the lexicon and is thus a principal feature of language structure is clearly different from the view taken in traditional generative linguistics, where formulaic language has been regarded as peripheral to a language (N. Ellis, 2008). For example, Lamb (1998: 169) supports the claim that linguists seem to underestimate the great capacity of the human mind to remember items while overestimating the extent to which humans process information by complex means of calculation rather than by simply using prefabricated units from memory.

2.3.2 Formulaic language in language processing

In the literature on formulaic language, there is converging evidence that formulaic language is processed both more quickly and potentially differently from non-formulaic language. This has challenged the existent model of language, specifically, the one designed by generative linguists. Wray (2002), for example, notes that formulaicity challenges the Chomskian claim that the language of normal adult native speakers is fully generated at the time of production and fully analysed in comprehension in two aspects: first, as a type of formulaic language, idioms cannot be processed in a generative way; second, as argued by Pawley and Sydney (1983), not all possible grammatical sentence occur with equal frequency or are judged to be equally idiomatic by native speaker.

Furthermore, Sinclair (1991: 110), for example, suggests that language as a whole is structured according to two principles: the open choice principle and the idiom

principle. The open choice principle is in line with the traditional view of language, that any word might be followed by any of a great number of words. The language user has almost a free choice; the only constraints being syntactic. The idiom principle brings about the selection of two or more words together, on the basis of their previous and regular occurrence together rather than at random. As for the relationship between the two principles, Sinclair stresses the non-random nature of language in his consideration of the need for the idiom principle and he states:

> It is clear that words do not occur at random in a text, and that the open-choice principle does not provide for substantial enough restraints on consecutive choices. We would not procedure normal text simply by operating the open-choice principle (ibid.).

He further explains the idiom principle is "put forward to account for the restraints that are not captured by the open-choice model" (ibid.). According to Barnbrook's (2009) review of Sinclair's work, the idiom principle is therefore suggested because of the inadequacy of the open choice principle to account for meaning in language. Thus, the user is provided with a large number of partly pre-constructed phrases that constitute single choices, even though they might appear to be analysable into segments (Sinclair, 1991: 110).

Inspired by this, Wray (1992, 2002) proposes dual-systems of both analytic and holistic language processing. Analytic processing entails the interaction of words and morphemes with grammatical rules, to create, and decode, novel, or potentially novel, linguistic material. Holistic processing relies on prefabricated strings stored in memory. The strategies preferred at any given moment depend on the demands of the material and on the communicative situation.

This is in line with Conklin and Schmitt's (2012) argument that our brains are likely to make use of a relatively abundant resource (long-term memory) to compensate for a relative lack in another (working memory) by storing frequently occurring formulaic

sequences. These can be easily retrieved and used without the need to compose them online through word selection and grammar sequencing (see also Pawley and Syder, 1983). Conklin and Schmitt (2012) also question whether native speakers are able to reduce demands on cognitive capacity because formulaic sequences are ready to go, are non-native speakers able to do the same? However, Foster's (2001) research provides evidence that non-native speakers seem to neglect phrases, focusing only on individual words[2].

In this section, related theories concerning formulaic language in language processing are reviewed. It can be concluded that formulaic language is a phenomenon presented in the minds of a language speaker. However, the issue of whether native speakers and non-native speakers share the same mechanism regarding formulaic language in language processing remains open. This will lead to research question 2 and 2a.

2.3.3 Formulaic language in communication

If we consider the role of formulaic language in facilitating communication, we can postulate a theoretical position that places formulaic sequences at the centre of language description, juxtaposing novelty and formulaicity as options for utterance construction (Wray, 2002:101). Liu and Huo (2011) summarise that formulaic language is important for communication in the following three ways:

First, according to Wray and Perkins (2000) and Wray (2002), in real-time communication, the use of formulaic sequences is double-edged in that it benefits both the speakers and the hearers. As to the speakers, formulaic sequences undoubtedly serve as a short-cut in the process of language production. By recalling familiar and prefabricated individual whole units, speakers do not bother to compose on-line through word selection and grammar sequencing, which, to a large extent, lessens the cognitive load and condenses the processing time. Once the brain is familiar with a

[2] More detailed discussion on formulaic language in language learning and use in the EFL context will be presented in the next chapter (Chapter 3).

linguistic task, it by-passes the normal processing route that was used to learn it, rather than just navigating it more quickly. In this way, the brain saves precious time of processing and is given much more time to attend to other information (McCrone, 1999). Such a saving in processing time seems to be valuable, particularly during demanding concurrent tasks (Wray, 1998).

Second, social interaction can be maintained by using formulaic language. Wray (1998) suggests that the reason why interactional functions are associated with formulaic sequences is because it makes them easier for the hearer to decode, which increases the likelihood of the interactional purpose being achieved. Formulaic language serves as a means to meet communication needs when language learners and users have inadequate linguistic competence. Hakuta (1974), suggests that formulaic speech occurs when people are forced to speak before they are ready. Nattinger and DeCarrico (2000) also agree that formulaic sequences enable learners to express functions that they are not yet able to construct from their linguistic system and thus promote fluency and ease motivational difficulties. More specifically, Wray and Perkins (2000) further identify three sub-categories of social interactional functions of formulaic sequences in interaction: 1) processing: it can save the speaker's processing effort; 2) interaction: it can manipulate the hearer (including the hearer's perception of the speaker's identity); 3) discourse marking: it can mark discourse structures (also see Wray, 2002).

Third, fluency is the ability to produce continuous speech without difficulties in comprehension or breakdowns in communication. Schmitt (2004) considers fluency as automatic procedural skill and processing power in using language.

The reviews in this section have highlighted that formulaic language is an important element in language grammar, language processing, and communication, rather than only a peripheral phenomenon. On the one hand, there is converging evidence from previous research on formulaic language that it provides certain advantages, notably speed, in language processing compared to non-formulaic language. However, arguably, the distinction between formulaic language and non-formulaic language is not clearly

demarcated but rather resembles a continuum, with formulaic and non-formulaic at two ends. On the other hand, when it comes to the issue of native speakers' and non-native speakers' formulaic language processing and use, it is considered difficult for second or foreign language learners to achieve native-like selection and native-like fluency. The issue of whether native speakers and non-native speakers share the same mechanism regarding formulaic language processing remains unclear.

2.4 Summary

This chapter intends to offer a reference point for the later analysis and so the review includes studies on formulaic language from different research positions, primarily from a corpus linguistics approach and a psycholinguistic approach. I established the rationale for this research by taking both speaker-external (focus on structure and function) and speaker-internal (focus on psycholinguistic salience in language users) approaches to identify and examine formulaic language in this research. These two approaches are realised by the methods used in corpus linguistics extraction of lexical clusters and analysis of psycholinguistic status, i.e., the intuition of language users. Then a review of the significant role of formulaic language in language grammar, language processing, and communication is provided.

I have argued here that as formulaic language varies across language users from different communities, it is not justifiable to use the intuition of others, especially native speakers' intuition, to judge non-native speakers' formulaic language. Therefore, the notion of formulaic language in this research is defined as 1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner, and/or retrieved with greater efficiency than other linguistic strings at the time of use, regardless of native language norms during identification; 2) Linguistic cluster, which meets a series of selective criteria in a corpus.

Chapter 3 The learning and use of formulaic language

In this chapter, the connections between studies of formulaic language, second language learning and second language writing will be established, to develop the rationale and framework for the current study. This chapter starts with a discussion of formulaic language and relevant theories in second language learning research. It will then elaborate on my reasons for choosing students' writing to study formulaic language in this study, which are also part of the rationale for this study. In the last section of this chapter, a review of previous studies on formulaic language in Chinese university students' written texts will be presented.

3.1 Formulaic language and second language learning

3.1.1 The role of formulaic language in second language learning

This section focuses on the role of formulaic language specifically in the second language learning process, including EFL and ESL contexts. First, the facilitating and complementary role of formulaic language in language learning approaches discovered by previous studies will be examined and justified. Second, this section will extend the view to a more general and ongoing picture of the position of formulaic language in second language learning. Therefore, some issues, like fossilization and L1 influence will be discussed in detail.

In section 2.3.2, Sinclair's (1991) language processing model is reviewed. Inspired by Sinclair's two principles (idiom principle and open choice principle), Wray (2002) promotes her dual-system solution to language processing: analytic and holistic processing (see section 2.3.2). She also claims that there are different processing mechanisms in L1 and L2 learners, as L2 learners only adopt an analytic approach. In other words, L2 learners can only use small elements (for example, words) to construct larger units, like phrases or sentences; or they only analyse the input into the smallest possible components and compose every utterance from these (Mauranen, 2007). However, there is evidence from previous studies suggesting that, in fact, second language learners may use two types of approaches at the same time or alternatively during second or foreign language learning,

as illustrated in the research conducted by Peters (1983), Skehan (1998) and Myles, et al. (1998).

Peters (1983), for example, concludes that there are two approaches to language learning existing at the same time: the gestalt approach and the analytic one-word-at-a-time approach. The gestalt approach mainly relies on holistic learning, which means learning to use the appropriate entire prefabricated language in different contexts by lexical units. Similarly, Ellis (1997: 13) suggests that learners must engage in both item learning and system learning. The former refers to learners internalizing chunks of language structure, usually considered as the basic units of language learning; the latter refers to their acquisition of rules, which means the knowledge that a given linguistic feature is used in a particular context with a particular function. Nevertheless, these two systems are interrelated during the language learning process.

Also, Skehan (1998: 54) proposes that there is a dynamic and interactive combination of the two systems in the mind's presentation and he notes

> Two systems coexist, the rule-based analytic on the one hand, and the formulaic exemplar-based on the other. In the former case, compact storage and powerful generative rules operate together to compute well-formed sentences. In the latter, the central role is occupied by a very large, redundantly structured memory system, and (presumably) less powerful rule which operate on chunks much of the time, rather than on individual items.

Skehan (1998) also explains that when time is available, learners are proficient at moving between the rule-based and the exemplar-based systems, and they do so naturally, but when time is in short supply, there may be a tendency for learners to rely on the system which is less labour-intensive and time-demanding.

In the current research, the second set of research questions (RQ 2 and 2a) aim to probe the role of formulaic language in the EFL learning from student perspectives. By answering these research questions, it is possible for the researcher to provide some

further evidence to address the issue of whether EFL learners have a different mechanism for language processing with regard to formulaic language from native speakers.

To address this issue, it is necessary to take into account some general observations on second language learning. First, it is inevitable to take the interlanguage into consideration when researching formulaic language in terms of second or foreign language learning. “Fossilization”, a term first introduced by Selinker (1972: 187) to refer to a permanent cessation in learning before the learner has attained target language norms at all levels of linguistic structure, needs to be taken into account. Based on this, Han (2004: 20) offers two explanations for fossilization. Thus, it can refer to both a product and a process. On a cognitive level, it involves cognitive processes or underlying mechanism that produce permanently stabilized interlanguage forms; on an empirical level, it involves those stabilized interlanguage forms that remain in learner speech or writing over time, no matter what the input or what the learner does. Ellis (2008), clarifies that fossilization is not an all-or-nothing phenomenon, which means that there is considerable variation in the linguistic form in the output of individual learners, as well as at different developmental stages. However, there are inadequate studies on formulaic language used by EFL learners which could be possibly raised from the fossilisation in interlanguage. As N. Ellis (2008) notes that despite the formulas being one of the hallmarks of child second language development. Still, advanced second language learners have great difficulty with native collocation and idiomaticity, producing many grammatical sentences which still sound unnatural and foreign.

Secondly, there has been little research on the relationship between the recurrent clusters and holistically stored formulaic sequences in the corpus. Schmitt and Carter (2004) mention the unspoken assumption that recurrent clusters identified by corpus analysis are also stored as holistic formulaic sequences in the mind. The argument continues that part of some clusters are also self-contained and stored holistically, like

idioms, but when it comes to lexical bundles, it is unsure for every learner that is all stored as a whole.

In corpus linguistics research this assumption has never been empirically put to the test, and thus, the extent to which recurrent clusters are psycholinguistically valid, or stored holistically, remains an open question. On the one hand, it is still problematic to define the exact relationship between the datasets, for example, particular learner corpora, and a more general characterization of interlanguage (Barlow, 2005). Barlow suggests that a learner corpus often represents just a single genre, and so some features of the learner's production may be closely associated with the genre rather than being more generally representative of their interlanguage (ibid.). It is sensible to assess the variability of different aspects of language production. Most studies on formulaic language within a corpus linguistics approach gather language production data at a single point, thus only providing a static description of language production, which is opposite to the viewpoint that the store of formulaic sequences are dynamic and is constantly changing to meet the need of the speaker (Wray, 2002). One the other hand, Conklin and Schmitt (2012) note that many of the corpus formulaic sequences were not stored holistically, but that this varies from individual to individual, and any particular formulaic sequence extracted from a corpus may or may not be stored holistically by any particular person.

The third issue related to formulaic language in second language learning concerns the influences of L1 and L2. In other words, the exposure time to the target language and the L1 should be accounted for. On one hand, Conklin and Schmitt (2012: 56) argue that although native speakers will automatically obtain the required exposure by adulthood, in many cases, non-native speakers will not, which explains why only relatively proficient non-native speakers, who have acquired their L2 over a long period, allowing them the time to amass sufficient language exposure, are able to process formulaic language in the quick and automatic manner of native speakers. Foster (2001) concludes that second language learners neglect phrases, focusing instead on individual words. On the other

hand, according to N. Ellis (2012), adults L2 learners already know about the existence of language units, categories and linguistic structure. Thus, the second language learners expect that there will be words and constructions in the L2 that correspond to such word classes and frames. N. Ellis also states that once the learners have identified them, or even once they have searched them out and actively learned such key vocabulary, they "are more like therefore to attempt creative construction, swapping these elements into corresponding slots in frames" (N Ellis, 2012: 40). However, the role of L1 has been neglected and underexplored for a long time in formulaic language research in second language learning, especially with the adult learners.

To sum up, this section has reviewed the role of formulaic language in second language learning research. It has discussed approaches to second language learning and use that distinguish between analytic and holistic processing. Also, three major issues regarding formulaic language in second language theories were revisited, including fossilization, the psycholinguistic validity of recurrent clusters, and influences from the L1. These aspects have generally been given less attention in the mainstream formulaic language research regarding second language learners, but are of great relevance for this study. In the next sections, the focus of the review will be on the language learner strategies used for formulaic language learning by second language learners.

3.1.2 Formulaic language learning

On the basis of language learning theories regarding the role of formulaic language discussed in the last section, we can note that formulaic language has been a disposable part of second language learning. This section will discuss the process of learning formulaic language in second language learning.

The previous studies indicate that there are processing advantages to using formulaic language, and the ability to rely on them is one factor that allows native speakers to be fluent. Schmitt (2000) holds the view that language ability requires not only the ability to produce language through syntactic generation (via grammatical competence), but also

the ability to use lexical chunks. He also points out that there are two types of lexical learning: explicit learning and incidental learning. Explicit learning focuses attention directly on the information to be learned, which gives the greatest chance for its acquisition; incidental learning can occur when one is using language for communicative purposes. He also concludes that the explicit learning is time-consuming and it could be laborious to learn an adequately sized lexicon; whereas incidental learning is slower and more gradual, lacking the focused attention of explicit learning (Schmitt 2000:120). The consensus is that, for second language learners, both explicit learning and incidental learning are necessary, and should be seen as complementary (Nation, 1995; Schmitt, 2000). Some studies report that both L1 and L2 writers may not only acquire formulaic language through formal instruction, but also through non-formal incidental learning, for example, extensive academic reading and repeated use of patterns through extensive writing (N. Ellis, 2008; Li and Schmitt, 2009).

Schmitt (2000) further relates language learning and processing to psychology by supporting the claim that in the case of vocabulary, the more one engages with a word, i.e. deep processing, the more likely the word will be remembered for later use; while, words not explicitly focused upon can be learned incidentally from exposure, facilitated by the use of vocabulary learning strategies.

As for the process of formulaic language learning, Peters (1983) proposes that learning lexical chunks is a three-part process: first, chunks are learned that are frozen wholes with no possible variation. At this point, they are unanalysed and are single lexeme, for example, some idioms, *like kick the bucket*, proverbs, and some expressions related to a functional use. Second, a language learner may realise that some variation is possible in certain chunks, which may contain open slots, until s/he realises that chunks can be segmented into smaller lexical units, often lexical words. At this stage, lexical chunks are partly fixed and partly creative. Last, the segmentation process may or may not continue until all the component words are recognised as individual units by use of syntactic analysis. Peters also gives a further explanation that there are some lexical chunks that

the learners may never start to analyse, and that may be retained only as unanalysed wholes. In addition, he suggests that much of a learners' vocabulary is learned in this way. This is especially because eventually learners are likely to eventually know numerous lexical chunks, seeing how they are easy to learn, efficient to use, and cover a wide range of lexical content (ibid.).

In this section, studies regarding two types of formulaic language learning: explicit and incidental learning have been reviewed. Also, I examined the learning process for lexical chunks by review study by Peters (1983). In relation to the research questions, this study is going to examine the processes of learning and using of formulaic language through accessing students' self-reports. It is hoped that this review will shed some light on the process of learning and use of formulaic language.

3.1.3 Strategies for formulaic language learning and use

This section will focus on the strategies used for formulaic language learning and use. It reviews related studies and theories on second language learner strategies first, and then narrows down the scope to the learning strategies that are employed in formulaic language learning and use specifically.

According to Rubin (2013), learning involves getting, storing, retrieving and using information. As such, many cognitive strategies are similar to processes identified by other second language learning researchers, that is, awareness, attention, hypothesis formation and testing, and practicing. Rubin indicates that learning strategies, considered inherent to the act of learning, are what students do when trying to learn and address their learning problem (ibid.).

The purpose of researching language learning strategies is to help learners to perform specific tasks and solve specific problems, to make learning easier, faster and more enjoyable and to compensate for deficits in learning (Cohen 2007: 38-39). Gu (2007) also notes that language learner strategy research focuses on the learners' decision-making process and the behaviours involving learning decisions aimed at maximizing results.

Cohen (2011: 682) provides a working definition of language learner strategies as below:

> language learner strategies can be defined as thoughts and actions, consciously selected by learners, to assist them in learning and using language in general, and in the completion of specific language tasks.

Cohen (2007, 2011) also offers classifications of language learner strategies, grouping them into three categories: strategies for learning and use, strategies according to skill area, and strategies according to function. The first category includes language learning strategies which relate to the learning of language material for the first time, and language use strategies, which relate to using material that has already been learned to some degree (Cohen and Weaver, 2006). Cohen (2011) clarifies that communication strategies (Cohen, 2004), i.e. a systematic attempt by the learner to express meaning by a target language in which the suitable systematic target language rules have not been formed, are also a type of language use strategies. Since learners may use them to avoid problematic language areas and express their meaning in some other way when they experience problems or breakdowns in communication. They might, for example, paraphrase words or coin words when they feel that communication is difficult.

The second category focuses on skill areas. This approach views strategies in terms of their role in listening, reading, speaking and writing. Also the strategies regarding vocabulary, grammar and translation are usually applied to all four of these basic skills areas.

The third category focuses on function, and includes metacognitive, cognitive, affective, and social strategies. Metacognitive strategies allow learners to control their language learning by planning what they will do, checking on progress, and then evaluating their performance in a given task. Cognitive strategies deal with the crucial elements of language use since they involve the processes that learners go through in both learning the target language (e.g., identification, grouping, retention, and storage of language material) and in using it (e.g., retrieval of language material, rehearsal, and

comprehension or production of words, phrases, and other elements of the target language). Social strategies encompass the means employed by learners for interacting with other learners and native speakers, such as asking questions to clarify social roles and relationships, asking for an explanation or verification, and cooperating with others in order to complete tasks. Finally, affective strategies help students regulate their emotions, motivation, and attitudes. In addition, they are used to reduce anxiety and provide self-encouragement (Chamot, 1987; Cohen, 2007, 2011; Oxford, 1990).

This section reviews some classic studies on second language learner strategies in relation to formulaic language learning and use, which provides the theoretical foundation for the description of the students' perceptions of formulaic language learning and use in RQ3. In this way, the learning and use of formulaic language by students can be organised and presented in a reasonable way.

3.2 Formulaic language and students' written texts

3.2.1 Why study written texts?

There are numbers of reasons that I chose the written texts as the medium to investigate the formulaic language learned and used by university students' in China. These reasons are the basis of a perspective that draws on both language learning and second language writing research.

3.2.1.1 The rediscovered role of writing in second language learning

The first reason is related to language learning. Traditionally, writing in a foreign language has been viewed within the learning-to-write perspective (Hyland, 2011). Within this perspective, writing is generally considered as the last skill to be learned because it is only when the L2 development is advanced that L2 writing can be effectively taught. It has also been considered to be the most distant reflection of the developing interlanguage, with the spontaneous oral language being seen as a much better approximation (Williams, 2012). As noted by Harklau (2002), L2 writing research is marginalised and the emphasis on spoken language is the result of the historical

development of the field of applied linguistics and parent disciplines of structural linguistics, linguistic anthropology, and child language development. Thus, writing is often seen as having a minor role in promoting second language learning development and a result of the acquisition.

Recently, the role of writing in second language learning has been re-framed with the writing-to-learn perspective gaining ground, which sees writing as a vehicle for learning (e.g. Harklau, 2002; Manchón, 2009; 2011; Williams, 2012). Williams (2012) suggests that writing because of its slow pace and the permanence of its record, may aid L2 development at several possible points: first, at the point of noticing and input processing, then, soon after the initial point of acquisition, as learners try out new and more complex forms or familiar forms in new contexts, and finally, later in the process, as learners retrieve and use forms over which they do not yet have full control, in other words, in knowledge internalisation, modification, and consolidation. The demand for precision and the opportunity to meet this demand stem from the two features that differentiate writing from other forms of language use. The permanent written record pushes learners to demand more of themselves regarding language form and the extended time gives learners the opportunities to meet this demand, often with the help of their explicit knowledge. Therefore, Williams claims that writing is not crucial to second language learning in all contexts, but has the potential to play a positive role, as writing provides the possibility of the conversion of explicit to implicit knowledge (ibid.).

The point of view expressed above supports the notion that writing could aid L2 learning and demands more from the learner's knowledge of L2. This can be seen as related to the feature of formulaic language as prefabricated in the process of learning, memorizing, and using, which is in accordance with the process of learning. Thus, it seems reasonable to explore formulaic language through written language, on the basis that the perspectives of learning-to-write and writing-to-learn are compatible and complementary in language learning, and writing is a typical medium employed in research to investigate language use in certain contexts.

There is one more reason why I have chosen to focus on written texts in this research. As one type of the samples of learner language, written samples are relatively permanent, and for this reason, easier to collect. In recent years, there have been large-scale projects on learner language based on written samples and motivated by the availability of computer-based concordance tools for analysis, for example, the International Corpus of Learner English (ICLE), the International Corpus of English (ICE). However, Ellis and Barkhuizen (2005) argue that various factors could have an impact on the linguistic characteristics of a written sample, and researchers need to provide careful descriptions of the data that have been collected. Such descriptions should, as a minimum, provide information about (1) the learners' social and situational backgrounds; (2) the situational context in which the writing took place; (3) the genre; (4) the topics; (5) timing; (6) availability of reference tools. In the next section, I will discuss research on formulaic language in students' written texts.

3.2.1.2 Conventional language use: formulaic language and genre

The second reason for examining formulaic language through written texts is that both formulaic language and genre studies focus on conventionalised language use. In this section, I will present the rationales that link the two areas.

A working definition of genre provided by Swales (1990: 58), has provided the foundation for later studies applying the concept of genre in linguistics and language education theory:

> A genre is a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains the choice of content and style. Communicative purpose is both a privileged criterion and one that operates to keep the scope of a genre as here conceived narrowly focused on comparable

> rhetorical action. In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience.

Askehave and Swales (2001) comment that this definition aims to set up a relationship between the purpose accomplished by a genre and the structure of the genre, and how the communicative purpose of a genre shapes the genre and provides it with an internal schematic structure. For example, in L2 writing research, Hyland (2009: 4) proposes his conceptualisation of genre regarding L2 research:

> It is based on the idea that members of a community usually have little difficulty in recognising similarities in the texts they use frequently and are able to draw on their repeated experiences with such texts to read, understand and perhaps write them relatively easily.

In other words, in linguistics research, genre research aims to "group texts according to types, and to identify and describe the features which texts of a particular genre have in common" (Caudery, 1998), and at the same time, to "represent how writers typically use language to respond to recurring situations" (Hyland, 2004:1).

J. Flowerdew (2002b: 23), notes that one of the important features of a generic view of language use is that studies regarding genre theory consider conventions as central to any form of generic description. Genres are essentially defined in terms of the use of language in conventionalised communicative settings, which give expression to a specific set of communicative goals of specialized disciplinary and social groups, which in turn establish relatively stable structural forms, and, to some extent, even constrain the use of lexico-grammatical resources. Bhatia (2004:23) summarised one of the common features of genre studies as:

> Genres are highly structured and conventionalised constructs, with constraints on allowable contributions not only in terms of the intentions, one would like to give expression to and the shape they often take, but also in term of lexico-

grammatical resources one can employ to give discoursal values to such formal features.

That is to say, the schematic structure of the texts and recurrent lexico-grammatical resources used in a discourse community are two features that can be used to describe the group practice of writing in academic or professional settings.

I proposed the following working definition of formulaic language earlier in this study: 1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner, and/or retrieved with greater efficiency than other linguistic strings at the time of use, regardless of native language norms during identification; 2) Linguistic cluster, which meets a series of selective criteria in a corpus, usually including the length of the cluster, threshold frequency, and text dispersion.

According to the second condition above, formulaic language has to be used and understood by members of a specific community, as it has to appear to be in several texts with minimum frequency. In this sense, formulaic language is essentially conventionalised language use, which it can be at the lexico-grammatical level, or the language used to achieve the schematic structure of the writing.

Hüttner (2005, 2007. 2008) extends genre analysis by examining two levels of conventionalised language use: she includes formulaic language as the micro-level of conventionalised language use and general analysis as (moves as genre constituents) the macro-level of conventionalised language use. The rationales behind her study are that while genre analysis and the study of formulaic language have been considered as separate areas of research, there are some overlaps as both research areas have focused on the conventionalised use of language, albeit at different levels (Hüttner, 2005). She points out that traditional formulaic language research mostly addresses the issues of phraseological conventions, whereas genre analysis studies are most likely to address the conventionalised structures of texts and their concurrent textualization

(ibid.). Her findings suggest that less proficient writers produce a higher frequency of formulaic language, and the reason for this is probably that inexperienced writers need to satisfy genre-specific demands and struggle to find appropriate words to complete the task, to fulfil their communicative intentions. Thus, formulaic language, in this case, can be seen as a building block which helps writers to create their texts (Hüttner, 2007).

However, the focus of the current research is to explore and describe the learning and use of formulaic language by EFL students in their written texts, which is not part of genre analysis studies. The idea of introducing genre studies and connecting it with formulaic language in this research is to extend views on formulaic language from the traditional approach (which usually focuses on sentences or lexico-grammatical level) to a schematic structure level.

3.2.1.3 Written text analysis: from language description to explanation

The third reason that I chose to study formulaic language through written texts, is because most previous research on formulaic language only provides linguistics description of the language products, or pure textual analysis, especially in written language (see section 2.2), rather than providing a thick description of the socio-cultural and psycholinguistic aspects. It is reasonable to rethink the approach to researching written texts in formulaic language studies.

Written text research has been viewed in different ways, for example, Flowerdew (2002a) provides a framework for the levels of language description in academic discourse, and presents three levels of language description, namely (1) language description as texts, (2) as genre, (3) as social practice. Bhatia (1993) views written texts as discourse and claims the research path has shifted from “discourse analysis as description” to “discourse analysis as explanation”, which means the analysis has gone beyond the linguistic aspects of text construction and interpretation to rationalize conventional aspects of genre construction and interpretation, also known as discourse

as genre. However, both studies present a view that extends textual analysis, and that considers the writer, reader, and the text to be in interacting with one another.

Bhatia (1993) suggests that linguistic analysis can be divided into three levels, including analysis of lexico-grammatical features, text-patterning or textualization and the structural interpretation of the text-genre. I will discuss each in turn.

In terms of the lexico-grammatical features research, Bhatia mentions that a text can be analysed quantitatively by studying the specific features of language that are predominately used in the variety to which the text belongs. This is usually done by undertaking a large-scale corpus linguistics statistical analysis of a representative sample of the variety in question (Bhatia, 1993: 24). However, depending only on this method tends to provide little information about the relation about the textualization and the way that communicative purposes are accomplished in a particular genre (ibid.).

At the level of text-patterning or textualization, Bhatia (1993) explains that it highlights the tactical aspect of conventional language use, specifying the ways in which members of a particular speech community assign restricted values to various aspects of language use (when operating in a particular genre. It provides information on form-function correlations by adding the functional explanation to the analysis of lexico-grammar in a genre. This is closely related to the corpus linguistic construct in the working definition of formulaic language that I proposed earlier: formulaic language is linguistic clusters and meets a series of selective criteria in a corpus. These criteria mainly contain the length of the clusters, the threshold frequency, and text dispersion. In other words, it specifies that formulaic language, as linguistic clusters, is recurrent and conventionalised language use within a discourse community.

Bhatia (1993: 11) further explains the language description as explanation and she notes that in order to introduce a thick description of language use, it is necessary to combine socio-cultural (including ethnographic) and psycholinguistic (including cognitive) aspects of text-construction and interpretation with linguistic insight, to answer questions, like

why are specific discourse-genres written and used by the specialist community the way they are. Further, Bhatia clarifies that

> text by itself is not a complete object possessing meaning on its own; it is to be regarded as an ongoing process of negotiation in the context of issues like social roles, group purposes, professional and organizational preference and prerequisites, and even cultural constraints (ibid.).

Here, I would like to argue that the language description as the text is not enough for the researcher to understand the language use in the community, and thus, elements outside the textual feature need to be taken into account.

This research aims to explore and describe formulaic language learning and use by EFL students in written texts. The first set of research questions (1a to 1d) focuses on the linguistic description of the form and function of formulaic language used by the students. The second set of research questions (2 and 2a) are concerned with psycholinguistic elements and the third research question is about the socio-cultural aspects relating to formulaic language learning, using and teaching at the research site.

3.2.2 Formulaic language in Chinese university students' written texts

The analysis of lexical clusters, as one type of formulaic language, has been long recognised for its important role in second language writing. Thus, Coxhead and Byrd (2007:134-135) conclude:

> 1) [lexical bundles] are often repeated and become a part of the structural material used by advanced writers, making the students' task easier because they work with ready-made sets of words rather than having to create each sentence word by word;
>
> 2) As a result of their frequent use, such [sequences] become defining markers of fluent writing and are important for the development of writing that fits the expectation of readers in academia;

> 3) These [sequences] often lie at the boundary between grammar and vocabulary; they are the lexicogrammatical underpinnings of a language so often revealed in corpus studies but much harder to see through analysis of individual texts or from a linguistic point of view that does not study language-in-use.

In accordance with the conclusion above, a substantial number of studies investigating formulaic language through corpora, especially written ones (Paquot and Granger, 2012). Mostly, these studies have investigated the use of formulaic language within English as a foreign language, and the level of proficiency targeted has more often been upper-intermediate to advanced learners; most studies have focused on learning writing, especially argumentative writings in the learner corpora (e.g. Paquot and Granger 2012; Nesselhauf, 2005; Waibel, 2008).

The Learner Corpora "are electronic collections of texts produced by foreign or second language learners" (Paquot and Granger 2012:30). Paquot and Granger (2012: 131) argue that there are two distinguishing advantages that make learner corpora "an ideal source of data to study the learner phrasicon". As learner corpora usually include contextualised language, rather than decontextualized words, phrases or sentences.

As for the corpora, many studies use large corpora, like the International Corpus of Learner English (ICLE), which is one of the largest and best-known learner corpora, and all texts have been collected for research purposes (Granger et al., 2002). However, studies based on the ICLE has produced a great number of studies on general non-native speakers' writings (e.g. Gilquin and Paquot, 2008; Granger, 1998; Paquot, 2010), but without a specific focus on the Chinese students.

There are some corpora containing some Chinese students' English written texts, for example, the BAWE (British Academic Written English Corpus). Chen (2009; also Chen and Baker, 2010) investigates the use of lexical clusters by Chinese, English writers in BAWE and professional academic writers. She, however, only examines Chinese undergraduate and masters level writing without distinguishing their levels study.

Besides, these texts were collected from the students who were studying in the UK and writing one text for the corpus. Therefore, it is very difficult to generalise the findings to all the Chinese EFL learners.

In addition to those corpora containing Chinese students' writing, a number of Chinese-only learner corpora have been compiled from university students in China, such as the WECCL 2.0 (The Written English Corpus of Chinese Learners) and the CLEC (the Chinese Learners English Corpus). The WECCL 2.0 contains 1.25 million tokens of times and un-timed argumentative and narrative essays written by English major students (Wen, et al, 2003). The CLEC has 1 million words of English compositions from senior secondary school students, English major students and non-English major university students (Gui and Yang, 2003). However, the access of these corpora is limited to the in-house use only, which is not with open access for public or research and studies based on these corpora mostly published in the Chinese language.

The majority corpus analytic studies on formulaic language in Chinese university students' writing are mainly within two paradigms: comparing formulaic language use in Chinese students' and native speakers' texts; or comparing formulaic language use by Chinese learners with different proficiency levels.

Research in the first paradigm usually applied a contrastive interlanguage analysis and the results generally converge: there are differences between Chinese university students' and native speakers in formulaic language use, and these differences were viewed as 'problems' by the deficit model of student writing. Typically, previous studies show the value of grouping lexical clusters according to their linguistic structures and/or functional use by the qualitative analysis. These structural characteristics and discourse functions of recurrent word sequences have been compared in the learner and native corpora, to identify errors and patterns of learner over- and under- use. Therefore, these differences need to be remedied or eliminated in order to achieve the nativelike attainment (Chen and Baker, 2010; Chen, 2009; De Cock and Granger, 2004; Granger, 1998). However, the L1 influence is often observed, but without further explanation. This

further leads to the discussion about the description of non-native speaker language production in the corpus with regard to the use of English as a lingua franca (ELF)[3]. Seidlhofer (2006) and Mauranen (2007) note that compiling the corpora to investigate language use among non-native speakers should be in order to offset the over-reliance on native speaker corpora.

As for the comparison in formulaic language use by Chinese student writer across different proficiency levels, a number of studies report that there is a positive correlation between the quantity of formulaic language use and the writers' English language proficiency level (or their curriculum stages) (Ding and Qi, 2005; Guo, 2011; Huang, 2014, 2015). Huang (2015) criticises this claim by arguing that 'quantitative gains is not paralleled by a qualitative gain' (Paquot and Granger, 2012:137), as many of these multi-word sequences are actually used inaccurately. For example, the lexical bundles investigated are of different lengths (from two to six words), and the settings used to extract them may vary considerably (Reppen, 2010; Chen and Baker, 2010). The issues discussed above lead to difficulties in directly comparing results among previous studies, however, there is a general trend showing that less proficient learners seem to be more reliant on lexical clusters, and this high degree of repetition may be due to their more limited lexical repertoire (Paquot and Granger, 2012; Hyland, 2012). In addition to this, studies in this strand are mostly based on quantitative analysis rather than qualitative analysis of the formulaic items (Huang, 2015).

In this research, the first main research question aims at describing formulaic language use by students from the Year 1 group and the Year 3 group, in order to establish whether there are differences in using formulaic language between learners with different proficiency levels. As reported in previous research on formulaic language, the relationship between formulaic language use and language proficiency level or writing expertise in academic writing production is unclear (Biber, et al, 2004; Chen and Baker,

[3] ELF (English as a lingua franca): English when defined as a lingua franca serves as a means of communication among non-native speakers in the "expanding circle" of world Englishes (Kachru, 1986) (see section 4.1).

2010; Cortes, 2004; 2008; Hyland, 2008a, 2008b; Juknevičienė, 2009; Leedham, 2011; Li and Schmitt, 2009; Myles and Cordier, 2016). This leads to research question 1d.

With respect to the few studies that are non-corpus analytic, they usually employed small scale case study (Li and Schmitt, 2009; Kinzley, 2011). For example, Li and Schmitt (2009) focused on the lexical chunks in the writing of a Chinese master student who was studying at a UK university, detailing the challenges in the student's use of chunks and examining her view of these changes. However, this type of case study is very hard to generalise further from findings as variables such as the particular course and institution and the particular individuals play a large part (Leedham, 2011:46). Therefore, in this study, I am going to examine formulaic language use by Chinese EFL learners from a medium-ranked university in China. It is hoped that the findings can be generalised to the contexts with a similar situation.

In conclusion, studies on the formulaic language of L2 English have yielded a wealth of interesting findings. However, Paquot and Granger (2012) note that some factors make it difficult to form a coherent picture. Besides the different sizes of corpora and the types of formulaic language investigated, the following factors lead to incompatibilities between the studies.

First, the influence of learners' first languages has been observed yet underestimated. Although, the importance of a learner's mother tongue has been illustrated, a number of transfer effects have not previously been well documented. Paquot and Granger (2012) suggest it is a field well worth exploring in more detail and the influence of L1 on learners' use of formulaic language may be more substantial than was previously thought, and may play a role in L2 learning that has not yet been comprehended. Therefore, the second research question in this study aims to ask students about the learning and use formulaic language from their own perspectives, rather than solely based on learner-corpus-based studies (also see section 3.1.1).

Second, there are only few longitudinal studies (Li and Schmitt, 2010), while most studies

are based on corpus data which was collected at a single point in time (e.g. Chen and Baker, 2010; Durrent and Schmitt, 2009; Gilquin, 2007). Therefore, studies that collect data over a period of time may be needed to represent learners' language production and development. In addition, studies based on corpus linguistics approach only focus on language production, there is not enough insightful information about the processes of learning and using formulaic language. Thus, in the current study, the students are also invited to talk about their own perspectives on English formulaic language learning and use. By examining the textual data and perceptual data, a more precise description of learners' formulaic language use and learning can be offered.

Third, Ellis and Barkhuizen (2005) categorise the data of learner language into three types, that is, non-linguistic performance data, samples of learner language, and reports from learners about their own learning. In the current research, the focus is on examining learners' textual data and perceptual data together. In other words, formulaic language can be researched through the samples of learner language, and at the same time, learners' perspectives can be explored through the follow-up interviews (see Chapter 5).

3.3 Summary

To sum up, in this chapter, the links between formulaic languages, second language learning and second language writing have been established. The focus of this study is to describe and explore the learning and use of formulaic language and student reflections on the teaching of formulaic language in an EFL context. This chapter presents the main reasons that I chose written texts as a medium to study formulaic language with EFL university students and reviews the previous formulaic language research conducted on students' written text.

The reasons that written texts are chosen as a medium to study formulaic language are as follows: First, writing plays an important role in second language learning, however, most previous second language learning research neglects this and put the writing as secondary data compared to the spoken language data. Second, through the concept of

genre, the similarities between formulaic language and genre studies are restated---- both areas of studies examine conventionalised language use in a discourse community. Third, departing from the idea of language description as explanation, this research will extend the description beyond the linguistic level, and include the socio-cultural and psycholinguistic elements that relates to the language use at the research site. By this means, it aims to provide a well-rounded description and tentative explanation of the learning, use and teaching formulaic language in the research context. Finally, a review of previous studies on formulaic language in student written texts is presented, including corpus-analytic and non-corpus studies, with focus primarily on Chinese university students' writing. In the next chapter, the research context will be examined.

Chapter 4 Research contexts: EFL in Chinese higher education

As suggested by both Bhatia (1993) and Swales (2004), it is necessary to include the situational contexts in the research as a part of a genre-text analysis. Bhatia (1993) notes that it is necessary to look at one's prior experience and background knowledge of the specialist discipline as well as the communicative conventions typically associated with it. Swales (2004: 4) also recommends including contexts (historical, social, material, and personal) in specific genre exemplars.

In line with this, this chapter aims to give a thorough view of the research site, which is a foreign language department of a university in China. In order to position this site, the discussion will start with a broad picture of the EFL situation in Chinese higher education and then focus on the local institutional context.

4.1 English learners and users in China

As the most popular language in the world, the global spread of English has affected all domains of human activity from language in education to international relations, therefore it is crucial to understand the role that English plays, the status that it has and the purposes it serves in different contexts (Kachru, 2011).Kachru's (1986) early work distinguishes three major types of users of English: (1) native users of English for whom English is the first language in almost all functions; (2) non-native users of English who use an institutionalised second language (L2) variety of English; (3) non-native users of English who consider English as a foreign language and use it in highly restricted domains. Kachru (1986) refers to speakers in the first group as members of the Inner Circle, the second group as members of the Outer Circle, and the last group as members of the Expanding Circle. In China, although there were 250-300 million people who spoke some English in 2006, English is primarily learned in the classroom as a foreign language and English is not used other than in the classroom (Yang, 2006).

This leads to another issue. In previous chapters, the terms ESL and EFL have been used interchangeably, for the reason that it is easy to unify the previous studies in writing

research across various contexts. Based on the definition proposed by Gass and Selinker (2008), “Second Language”, as umbrella term, refers to the language that is learned after a native language and it can mean both learning a language in a classroom, as well as a situation of more “natural” exposure (Gass and Selinker 2008: 4). Specifically, EFL is learned by people who already use at least one other language and who live in a community in which English is not normally used. Such a community is inevitably influenced by norms that are not those of English-speaking countries, and those norms influence the teachers’ and learners’ expectations of the language learning process.

Tomlinson (2005) argues that most learners of EFL learn English in school together with a large group of peers of similar age and proficiency. They typically have a course book, are preparing for an examination, and are taught by a teacher who is not a native speaker of English. However, Tomlinson argues, there are many variations on the EFL scenario, for example, the size of the class can vary from 10-100 students; the different equipment in the classrooms, different materials, and the teacher might not be skilled, professionally educated or experienced. Whatever the social context of learning, it will affect the learners’ expectations and behaviours.

4.2 English major degrees in Chinese higher education

According to the latest English Curriculum Standards for English Majors in Higher Education produced by the Committee of Higher Education English Language Teaching (CMELT, 2000), the English major in higher education aims to equip students with solid English language based comprehensive cultural knowledge, and interdisciplinary skills using English such as foreign affairs, education, trade, technology, and military, to conduct translation, teaching, management, and research.

The Curriculum Standards also decides on the requirement for the curriculum design of the English major for English BA degree. Generally, there are two levels over four years of study for the degree programme: the junior years (including year 1 and year 2,) and the senior years (including year 3 and year 4). The main tasks for the junior years in

English major teaching are to lay the foundation of English language knowledge and train students in basic skills (i.e. listening, reading, writing, speaking and translating) in the English language and to improve their ability to use English in practice. In the senior years, students are directed to acquire a more specific knowledge in English and to improve their sense of cultural differences, as well as their communicative skills in English.

The courses in the English major programme can be grouped into three categories: English language skills, English courses and specific courses in English. The first category, English language skills, refers to the integrated training and individual skills training, such as integrated English, listening, speaking, reading, writing, interpreting, translating, and so forth. English courses include modules related to the English language, culture, and literature, e.g. linguistics, lexicography, English grammar, English literature and Western cultural studies. Specific courses in English are mainly modules from other disciplines that are related to the English language, for example, foreign affairs, international business, management, education, or military studies. Based on this requirement, the total learning hours for the four years are no less than 2,000 hours, excluding optional modules, and no more than 2,200 hours. Also, the syllabus can be designed locally, according to specific conditions in each university. The curriculum also provides guidelines for each band (each semester is considered as one band, for example, band one is the first semester, band eight is the last semester of the fourth year), with regard to pronunciation, grammar, vocabulary, listening, speaking, reading, writing, translation, dictionary use and cultural knowledge.

In this study, I am going to investigate the written texts produced by students from Year 1 and Year 3, who are supposed to have achieved the requirements of Band Two and Band Six. In this scheme, with reference to the writing requirement guidelines, students in Band Two are expected to be able to write a 100-word text clearly, without severe grammatical errors, paraphrase some paragraphs in the textbook, write a short letter, a notice, and notes in the correct formats, with acceptable handwriting. In Band Six,

students are asked to write the summary of a story, a book review, and reports or formal letters in English. Emphasis is put on using English correctly and appropriately, and students are expected to write about 250-300 words in thirty minutes.

In the section on Teaching Principles, the Curriculum Standards re-state that the primary task of English language teaching is mastery of the basic language skills, that is the listening, speaking, reading, writing and translation skills, which are supposed to be learned during the four years of study. Among these five basic skills, more attention should be paid to the speaking, writing and translating skills. Teachers should also guide students to a greater awareness of cultural differences and foster the ability to respond to cultural differences in tolerant and flexible ways. The Curriculum Standards requests students to undergo TEM-4 (Test for English Major Band 4) and TEM -8 (Test for English Major Band 8) during the second semester of the second and fourth year respectively.

According to Dai and Zhang (2007), there are already more than 900 universities in China running a BA English programme, and since 2004, the number of candidates taking TEM 4 and TEM 8 has increased by 10% per year. In order to maintain and improve the quality of the English major degrees, the MoE started to assess all the BA English programmes in China's higher education system in 2005 (Dai and Zhang, 2005; 2007). There are some reports based on the assessment results with regard to the issues such as the mismatch between the guidelines of the Curriculum Standards and university syllabus design, lack of hours in English language teaching, and weaknesses in students' basic language skills (Dai and Zhang, 2007; Hu and Sun, 2006; Zhong, 2006).

Based on the curriculum description, it is clear that the English major programmes in China combine language and content learning. On the one hand, students need to learn English language during the first two years to acquire further language skills on the basis of secondary school English education. In this way, students have to make the adjustment from studying all subjects in their national language, Mandarin, at school level to studying subjects in English at tertiary level. On the other hand, in the last two years at the university, students have to learn study content-based courses in English. In

line with studies by Dudley-Evans (1999), and Dudley-Evans's and St John (1998), the subject lecturers, most of whom are themselves non-native speakers of English, may deliver lectures in a mixture of English and the national language, in this case, Mandarin. However, notes are dictated in English, and examinations and assignments are written in English, which usually do not usually make heavy linguistic demands on students. Students are tested through multiple choice questions, and required to write short texts in exam papers.

Within the Chinese context of English language teaching and learning, several issues have affected the interest of applied linguistic researchers (Jin and Cortazzi, 2002; Zhu, 2003). The first is the demand placed on the teachers who are non-native English speakers. Although there are increasing numbers of native English teachers available in higher education in China, the majority of English language teachers are Chinese and have never been outside China. Owing to a lack of English proficiency, some Chinese EFL teachers are concerned about not being able to answer spontaneous questions about the target language, sociolinguistics, or culture as they arise from interactions in the classroom (Rao, 2002).

Secondly, EFL teaching is considered to be difficult to conduct in a communicative way in the classroom in China. The traditional relationship between Chinese teachers and students is characterised by hierarchy (Chang and Holt, 1994; Wang and Cortazzi, 2013). This makes it unfeasible to suspend their beliefs about the teacher's role in the classroom and use a more student-centred way of learning. Communicative Language Teaching (CLT) treats language as communication, rather than as code learning, and so is different to more traditional approaches, like audio-lingualism, which are popular in China. CLT explores "the possibility of non-analytical, participatory, or experiential ways of language learning as a deliberate teaching strategy" (Stern, 1983: 473). Although CLT was introduced into China in English language education in the early 1980s, its success has not been universal, and there have been criticisms especially in public sector (Hu, 2005; Liao, 2004; Yu, 2001). For example, based on their own teaching experiences with

Asian students, Ballard and Clanchy (1991) describe Chinese students as quiet, respectful of teachers and textbooks, and reluctant to ask questions or express their own opinions publicly. The authors think that these students learn by imitating others rather than by independent thinking; their learning, therefore, is seen as reproductive rather than analytical or speculative. In order to get a higher score in examinations, students prefer to attend lectures on intensive reading and grammar (Rao, 2013), which is also in accordance with my own experience as a student in the context at that time.

The third problem lies in the more general educational culture in China, which is heavily examination driven. The modern examination system in China evolved indirectly from the old imperial examination system which, in the prior two millennia, helped to select the best administrative officials and maintained the peace of the country. Examinations, still play a pivotal role in student success in China, as China has a highly selective educational system with fewer students at the higher end of the educational ladder and students face numerous examinations as soon as they start their schooling. Only top-performing students in high-stake tests may enrol at prestigious schools and universities, get recruited for satisfactory jobs, and enjoy opportunities that are off-limits to lower scoring students (Kirkpatrick and Zang, 2011).

Additionally, the lack of resources and equipment is a consideration in current EFL development in China. In most schools, there are 60-80 students per English class. The schools or universities in the western provinces or suburbs, for instance, do not have enough equipment, e.g. audio-visual, or internet access, which can negatively affect the learning experience of the students.

4.3 Research site description

The fieldwork for this research took place at a state-owned university in China, Hubei University of Automotive Technology. As I consider the context for this research to encompass both Higher Education in China more generally, and the English major degree programme at the Hubei University of Automotive Technology, I will provide information on these two aspects in turn.

4.3.1 Higher education in China

The reasons for choosing Hubei University of Automotive Technology as my research site are as follows:

Of the 2,305 universities authorized in China currently, 1,649 are public. Of these, 111 are sponsored by the national central ministries and agencies directly and are thus widely seen as elite or top-tier universities, for example the Hubei University of Economics, Dalian University of Foreign Languages, Sichuan International Studies University. The remaining 1538 universities are run by local authorities, (MoE, 2013). In China, it is commonly considered that public universities are better than private universities. Universities in China generally select their students based on their performance in the National Higher Education Entrance Examination. The entrance scores required by public universities are typically much higher than those of private ones. The reason for this is seen in the fact that private universities are a recent phenomenon and are seen as academically less competitive. Li et al. (2012) note that since educational resources are highly limited in developing countries such as China, governments are compelled to make a trade-off between improving university quality and expanding access to higher education, for example, through allowing private universities. Students and their parents also devote considerable effort and resources to gaining entry to high-ranking universities.

The Hubei University of Automotive Technology is a medium-ranked, public university, run by the local authority (in this case Hubei provincial government). It thus seems to represent an average Chinese university, which is not a privileged top-tier institution, but is still locally prestigious and academically demanding.

A further issue to consider in China, the world's largest country, is the geographic location of any educational institution. Thus, it is sensible to take the economic and educational condition into account when considering issues of English language learning and teaching in China. Hu (2005) proposes that the differences in English education between developed areas and underdeveloped areas in China can be attributed to the joint influences of a range of social, economic and cultural factors. Similarly, MacKay

(2012) states that one of the major causes of the inequality of access to English learning is the economic divide, which is often reinforced by the Minister of Education and China is a case in point. Thus, a minority of elite schools or universities can take advantage of their much greater volume of various types of capital – greater financial resources, excellent infrastructure, wider social networks, well-trained staff, and high calibre students.

The economic reforms in China in the late 1970s, saw an abandonment of the Cultural Revolution and the government launched a national modernization program in which English education was seen as a key component. Part of this initiative was the introduction of compulsory EFL learning at primary 3, i.e. ages 9 to 10. The directive also mandated that efforts in promoting English language proficiency were to be aimed at strengthening English language teaching in elite schools, which were expected to produce the English-proficient personnel needed to successfully undertake national modernisation. In fact, in 1985, the Minister of Education gave several economically developed provinces and municipalities the autonomy to develop their own English curricula, syllabi, and textbooks for primary and secondary education (Mackay and Rubdy, 2009:1992). These materials tended to be more innovative, learner-centred and communicative than earlier classroom texts and materials. Cortazzi and Jin (1996) comment that there continue to be significant differences in language teaching developments between the major cities and small cities, between rural towns and countryside, between coastal and inland areas, between north and south and between key and non-key schools or universities. Hu (2005) recognises that the gap in the quality of English language education is one of the major issues in current English language teaching in China. Since the increasing regional socioeconomic disparities and their influences on educational development are recognized, different educational requirements have been set for different regions.

The participant university, Hubei University of Automotive Technology, is located in Shiyan City, which is at the juncture between the more developed regions of eastern China and developing western China. Shiyan is a third-tier municipal city with around 3.34 million inhabitants. People there speak the local dialect, which is a variety of

Mandarin. The city itself is an important centre for the automobile industry in China, with Dongfeng, the premier Chinese truck, bus, and heavy goods vehicle as a major employer. It is also labelled as the “Detroit” of China, given this focus on automobile manufacturer. In this case, with regard to the implementation of English education policies, English language teaching and learning, in this university is neither radical nor conservative.

The university chosen for this research gives priority to engineering in the education of the undergraduate and graduate students, especially in automobile, mechanics, materials and electronics, compared to other disciplines such as management, economics, humanities, social sciences and law. The majority of the students aim to find a job in the automobile industry or related to automobile market after graduation.

4.3.2 Local context of the university

Similar to English departments in other Chinese universities, students are required to take the National Matriculation Test (NMT, also known as Gaokao in Mandarin) in order to enter the Hubei University of Automotive Technology. The NMT (Gaokao) involves five or six subjects depending on the requirements of the type of university to which the student applies. Chinese, Mathematics, and English are three compulsory subjects for all candidates regardless of their choices of the university (Cheng, 2008). All universities are categorised into different tiers and those belonging to a higher tier are afforded first priority in admitting students. In all provinces, students fill in a form, on which they list their university preferences (4-6 in each tier) and favoured majors in order of priority. The timing of turning in the preference form varies by province: it may be submitted prior to the exam, after the exam but prior to being informed of the scores, or after learning the scores. For the admissions procedure in most provinces, in the first round, each university considers only the students who list it as their first choice. An applicant with a total NMT score above its threshold score is accepted; an applicant with a score below the threshold is rejected and placed in a pool of candidates for the university that is next on the list of preferences. Given the shortage of high-quality institutions, students have little chance for admission by their second choice universities if they fail to enter the first choice university. Therefore, filling in the preference form requires

certain skills and strategies, since two students with the same NME scores may end up in very different universities. Once the student is admitted to a university, the process is terminated for this student, and s/he can no longer change their universities. Through this mechanism, a student obtains at most, one offer. If the student turns down the offer, it means she or he will not be going to a Chinese university that year.

The Hubei University of Automotive Technology is placed in the second tier and is medium ranked among all the Chinese universities. The department of Foreign Language, where the participants come from, was founded in 2001, and hosts only one four-year programme in English, BA English Automotive Trading, and Business English. Currently, the department has 14 classes with around 400 enrolled students. The students are admitted to the current programme through the university admission mechanism described above.

During their four-year degree, the students are required to attain a good command of listening, speaking, reading and writing and translation skills in English, a good comprehension of international trade and a general knowledge of automobile information (Zou, 2003). Courses related to English language skill training are mainly set up for them in the first two years at university, and include Integrated Skills of English, Extensive Reading, A Basic Course in English Writing, Pronunciation, Oral English, Advanced English Grammar, Intensive Listening, Translation Theories and Practice, Extensive listening, Comprehensive, Contemporary College English. From Year 3, some English courses and a specific course in English are added into the syllabuses, such as English Literatures, Linguistics, Interpretation, English Culture, Automotive Engineering English, and International Business. In general, these courses are in accordance with the Curriculum Standards. Additionally, some courses on automobile and international business are also taught in both Chinese and English.

For English majors in Chinese universities, the Test for English Majors (TEM) is an important test. It assesses the language proficiency of English majors and is administrated by the National Advisory Commission on Foreign Language Teaching in Higher Education (NACEFL) in China. Another purpose of the test is to promote English teaching and learning for English majors (Cheng, 2008). TEM is a criterion-referenced

test, where students' performance is evaluated against the criteria stipulated by the teaching syllabus (Zou, 2003). The tests consist of two levels: TEM-4 administered at the end of the second year and TEM-8 at the end of the fourth year of the undergraduate program. Students on this programme can only be awarded a degree if they pass the TEM-4.

The total number of staff in the English department at the Hubei University of Automotive Technology is 54, 6 of them are foreign teachers, 6 are administrators, 6 teach other foreign languages (mainly German and Japanese) and 34 Chinese nationals hold full-time EFL teaching positions as professors, lecturers or teaching fellows. The Hubei University of Automotive Technology is not an exception within the broader context of EFL in China, in sharing a text-focused and teacher-centred tradition of learning English with rote-learning and a strong emphasis on exams at the university level.

For example, many teachers use passages from the textbooks in their PowerPoint presentations directly as teaching materials in the classroom. Basically, they pay much attention to and analyse the grammar underlying any text during teaching. The teachers also choose the key content for teaching from the textbooks. For instance, in the first class of A Basic Course in English Writing, the teacher listed the key points for teaching on the basis of the contents page and decided which parts were important since they may be examined in the final exam of the semester.

As for the English language learning at this university, it is dominated by rote-learning and strongly exam-oriented language learning. Pervasively, students work hard on memorizing vocabulary and grammar rules, as well as attending more drilling training sessions on specific language skills for the various tests and exams after class. For example, if they are preparing for the final exams of the semester, they try to memorize the content and exercises from the textbook, which will probably appear in the exam paper in their original form without any changes to the textbook. While they are preparing for the specific tests, such as TEM-4 or TEM-8, intensive drill training sessions play a vital role. These exams are designed for English majors, and every candidate only can take the exam twice. If they fail the test both times, they will have no chance to

make up. Therefore, all students in the department work hard to pass the tests, especially TEM-4, which is one of the conditions that students have to meet before they can be awarded a BA English degree.

In such a context, there is a contradiction in the empowerment of teachers and students. It is difficult to say who is more empowered. For example, the teachers seem to have the authority in the classroom, they have right to decide which content is taught or included in the final exams. In terms of conduct, teachers also may decide, thus, they make it clear that mobile phones are absolutely not allowed in the classroom, but it is not unusual if a teacher's phone buzzes in the lecture.

As for the assessment, the students are given the final grade each semester through an online credit system. The grades consist of an exam grade and a semester grade in a ratio of 7:3. The semester grade consists of various components, e.g., homework, class assignments, and attendance. The assessment during each semester is formative, and the teachers give feedback on homework that students submit. Homework appears to have various functions, e.g. attendance checking, exam preparation. The final exams in each semester are mostly summative.

4.4 Summary

This chapter has provided some situational information on the student writers, and the research site by narrowing down scope step by step. At the beginning of this chapter, I drew on Kachru's circles of English language users to position these learners as Expanding Circle English users. A general picture of English majors in Chinese higher education was presented by examining the curriculum, the actual situation in language teaching and learning, and some issues and concern raised by researchers which are pertinent to the case of China. Led by this, a further description of a local institution is delineated at a general and local level in the context of the Hubei University of Automotive Technology, my research site. As a medium-ranked university, it can to some extent be considered a typical case among other universities in China.

Chapter 5 Methodology

In the preceding chapters, previous research on formulaic language, and its relation to second language research and second language writing research has been discussed. Furthermore, the research site has been situated by reviewing EFL in the Chinese higher education and the local context. On account of this, the rationale of this study, i.e. exploring formulaic language through student written texts has been further established and justified. In this chapter, the methodology of this study is going to be presented, which forms the basis for operationalization in this research.

5.1 Research questions

As discussed in section 1.2, there are three main research questions with sub-questions to guide the present research. These are:

(1) To what extent do Chinese university students use formulaic language in their written English?

a) What are the main structures of formulaic language used by these learners?

b) What are the main discourse functions of formulaic language used by these learners?

c) What is the relationship between the distribution of structural and functional categories of formulaic language in the learners' written texts?

d) How is formulaic language used differently in the written texts of Year1 and Year 3 university students?

(2) What do Chinese university students perceive formulaic language to be?

a) To what extent is this perception different in Year 1 and Year 3 students?

(3) How do Chinese university students perceive the learning, use and teaching of formulaic language?

These research questions are designed to shed light on the use, learning, and teaching of formulaic language in the Chinese EFL context to help understand students' conceptualisation of formulaic language, as well as their learning, use and the teaching of formulaic language at the research site. The first set of research questions (RQ1a to 1d) aims to explore the linguistic structures and discourse functions of formulaic language in Chinese EFL learners' written texts. In order to provide a thorough description of formulaic language use by the students in their written texts, I will compare the texts from two student cohorts, i.e. students from Year 1 and Year 3 by two identification approaches: corpus-extracted linguistic clusters and student identified formulaic strings. The relationship between the distribution of structural and functional categories of formulaic language in students' written texts will also be investigated. The second set of research questions (RQ2 and 2a) aims to establish the students' conceptualisation of formulaic language and examines if there is any difference in the perceptions of the two student cohorts. The third research question (RQ 3) aims to reveal the sources, processes, and strategies that the learners employed in formulaic language learning and use, as well as their comments on formulaic language teaching.

The answers to these questions should further develop the current theoretical and descriptive framework of formulaic language and its relation to learning, use, and teaching in EFL contexts. In terms of research methods, the first set of research questions focuses on researching students' written texts, and will employ a textual analysis based on the corpus linguistics approach and language users' perceptions. Therefore, the methods are combined with quantitative and qualitative analysis. RQ 2 and RQ 3 focus on the student writers, and aim to depict their perspectives on the phenomenon of formulaic language, and are thus exploratory and qualitative.

5.2 Research design

5.2.1 Research methods: mixed methods

Research methods in Applied Linguistics, which many researchers position at the interface in Social Science and Educational research, employ a varied inventory of methodologies with specialisations (e.g. Cohen et al., 2011; Fraenkel and Wallen, 2006; Larsen-Freeman and Long, 1991; Lichtman, 2012; Nunan, 1992; Suter, 2011).

In this section, I will provide the rationales behind the research design used for this study. Therefore, first I will examine the attributes of the qualitative and quantitative research methods respectively. I then discuss the "paradigms war" between the quantitative and quantitative research approaches. This introduces a discussion of the model of mixed methods research design, which leads to the specific models that I have used to solve each research question for this study.

5.2.1.1 Qualitative and quantitative research methods

In the early days, researchers in Applied Linguistics explored different research methods (Seliger and Shohamy, 1989). Johnson (1992) clarifies correlational approaches, case-study approaches, survey research, ethnographic research, experimental research, and multisite/multi-methods research. Nunan (1992) explains some research methods in detail, such as experimental methods, ethnography, case study, classroom observation research, introspection methods, elicitation techniques, interaction analysis and programme evaluation. McDonough and McDonough (1997) elaborate on observations, introspections, dairies, experiments, interviews, questionnaires, numerical techniques, and case studies. Brown and Rodgers (2002) give specifications for case study research, introspection research, correlational research, quasi-experimental research, and course or programme evaluation.

In general, the key distinction between the different types of research is that between qualitative and quantitative research (Brown 2004: 486). Usually the two methods are

labelled and used interchangeably by different researchers: i.e. quantitative and qualitative research (Brown and Rodgers, 2002; Mackey and Gass, 2005) to describe different methods of data collection, such as psychometric and naturalistic research traditions (Nunan and Bailey, 2009), normative and interpretive research (McDonough and McDonough, 1997); analytical-nomological and exploratory-interpretive (Grotjahn, 1987); confirmatory and descriptive research (Ellis, 1997). Dörnyei (2007: 24) proposes a primary definition and typical examples of the two approaches:

> Quantitative research involves data collection procedures that result primarily in numerical data which is then analysed primarily by statistical methods. Typical example: survey research using a questionnaire, analysed by statistical software such as SPSS.
>
> Qualitative research involves data collection procedures that result primarily in open-ended, non-numerical data which is then analysed primarily by non-statistical methods. Typical example: interview research, with the transcribed recordings analysed by qualitative content analysis.

Traditionally, many studies on research traditions have made a binary distinction between qualitative and quantitative research. For example, Reichardt and Cook (1979) conclude the terms commonly associated with quantitative and qualitative approaches to research. They provide a conclusive comparison, including the research methods, the researcher's role, data collection, data analysis and researcher's orientations, as illustrated in the figure below:

Qualitative Paradigms:	Quantitative Paradigms:
• Advocates the use of qualitative methods; • Concerned with understanding human behaviour from the actor's own frame of reference; • Naturalistic and uncontrolled observations; • Subjective; • Close to the data; the "insider" perspective; • Grounded, discovery-oriented, exploratory, expansionist, descriptive, and inductive; • Process-oriented; • Valid: real, rich, and deep data; • Ungeneralizable; single case studies; • Assumes a dynamic reality	• Advocates the use of quantitative methods; • Seeks the facts or causes of social phenomena without regard for the subjective states of individuals; • Obtrusive and controlled measurement; • Objective; • Remove from the data; the "outsider" perspective; • Ungrounded, verification-oriented, confirmatory, reductionist, inferential, and hypothetical-deductive; • Outcome-oriented; • Reliable, hard and replicable data; • Generalizable; multiple case studies; • Assumes a stable reality

Figure 5.1: Binary distinctions between qualitative and quantitative research in Reichardt and Cook (1979: 232)

Dörnyei (2007:39-41) synthesises the strengths and weaknesses of quantitative and qualitative approaches in the following ways: for qualitative research, he proposes that the significant strengths are: it is an effective way of exploring new, uncharted areas, making sense of complexity, answering "why" questions, broadening our understanding of human experience, longitudinal examination of dynamic phenomena, flexibility when things go wrong, and providing rich material for the research report. The weaknesses of the qualitative research are summarised as: small sample size and difficulty of generalizability, researchers' role, lack of methodological rigour, theories that are too complex or too narrow, and it is time-consuming and labour-intensive (ibid.).

In the field of the quantitative research, Dörnyei (2007:35) presents the following presumed advantages; it is systematic, rigorous, focused and tightly controlled, involving precise measurement and producing reliable and replicable data that is generalizable to other contexts. As the downside of quantitative research, Dörnyei points out that on one hand, it averages out responses across the whole observed group of participants, and by working with concepts of averages it is impossible to do justice to the subjective variety

of an individual life. In particular, similar scores can result from quite different underlying processes; on the other hand, quantitative methods are generally not very sensitive in uncovering the reasons for particular observations or the dynamics underlying the examined situation or phenomenon (ibid.). Based on these considerations, Dörnyei (2007) sets up a framework of mixed methods in Applied Linguistics.

In this study, I will use the most popular employed terms quantitative and qualitative research methods to describe the two ends of the qualitative-quantitative continuum. In terms of formulaic language studies and writing research, most research done within a quantitative approach only depicts the language production (for example, text analysis, corpus analysis, discourse analysis), but with little information about the language learners' viewpoints in written texts according to different contexts. This has led Lillis (2008:353) to propose that strategies are needed for "closing the gap between text and context". Specific to the corpus-based approach, Flowerdew (2005) claims that corpus studies lead to the atomised, bottom-up description of language use, and so a corpus study does not consider the contextual aspects of texts. Harwood (2006) provides advice on how contextual features can be incorporated into corpus studies, for example, by using interviews and focus group discussions with users in or order to obtain insightful information about the corpus.

As can been seen in the research questions, the first set of research questions require numerical information (RQ1a to 1d); while the other two sets focus mainly on exploring the reasons underlying the phenomenon (RQ2, 2a and 3). It is sensible to argue that these research questions cannot be solved by either the quantitative or qualitative paradigm only. Therefore, this study is going to employ mixed methods to seek the solutions to the research questions.

5.2.1.2 Mixed methods research

In response to the limitations of the sole use of the quantitative or qualitative method, mixed methods research emerged and is considered a legitimate alternative to these two traditions (Talmy and Richards, 2011). In other words, mixed methods research employs a variety of approaches to answer the research questions that cannot be answered by using a single method (Doyle, et al., 2009). Newman and Benz (1998) also argue that quantitative and qualitative approaches should not be viewed as polar opposites or dichotomies; instead, they represent different ends of a continuum.

Researchers, who hold that there is a chasm between quantitative and qualitative research methods, perceive quantitative and qualitative research as distinct and based on fundamentally different principles (Bryman, 2007). Quantitative methods are within the positivist scientific model, which works on the basis that there is a single reality and therefore seeks to identify causal relationships through objective measurement and quantitative analysis (Firestone, 1987). In the paradigm of positivism, the researcher is considered independent and objective, using larger samples to test carefully constructed hypotheses. The prevailing tenet is that the researcher in the positivist tradition can avoid bias in a process of inquiry (Creswell, 2009).

However, constructivism also described as one of the interpretative approaches (Creswell, 1994), guides qualitative researchers to examine the context of human experience (Schwandt, 2000). Doyle et al. (2009) conclude that there are multiple realities and so different interpretations may result from any research endeavour. Thus, in the interpretative paradigm, the researcher is subjective with the focus directed at the deeper understanding of what is happening with a smaller group of people.

Researchers like Creswell et al. (2008), Dörnyei (2007), Grotjahn (1987), and Nunan (1992), have argued that the qualitative/quantitative distinction is an oversimplification, and it is necessary to take into consideration the methods of data collection, the type of data yielded by the investigation, and the type of data analysis.

Combining quantitative and qualitative methods in a study has emerged and regarded as "the third research paradigm" (Bryman, 2007), "the third methodological movement" (Teddlie and Tashakkori, 2009), or "the third path" (Gorard and Taylor, 2004). The philosophy underpinning mixed methods is pragmatism (Morgan, 1998, 2007; Tashakkori and Teddlie, 1998; Patton, 1990), which conveys its importance of focusing attention on the research problems in social science research and then using pluralistic approaches to derive knowledge about the problems (Creswell, 2009).

It is called variously by different researchers: "multi-strategy" (Bryman, 2004), "mixed methods" (Creswell, 2003; Dörnyei, 2007; Tashakkori and Teddlie, 2003), "hybrid research" (Ellis, 1997), or "mixed methodology" (Tashakkori and Teddlie, 1998). Tashakkori and Creswell (2007: 4) define mixed methods research as

> where the investigator collects and analyses data investigates the findings and draws inferences using both qualitative and quantitative approaches or methods in a single study.

In light of this, "mixed methods" is used to describe the research method employed in this study, as both quantitative and qualitative methods are employed in the study, as well as naturalistic data (e.g. students' written texts) and the responses elicited from interviews.

Moran-Ellis et al. (2006), for example, clarify the distinction between integration and combining methods in mixed methods. They argue that integration requires that different methods (or type of data) are given equal weight, and, with regard to operationalization, that they are orientated to a common goal or research question and therefore are necessarily interdependent while retaining their paradigmatic modalities. Combined methods, on the other hand, means that the qualitative component is an adjunct to the quantitative, improving its depth or quality, rather than positioned as making an equal contribution to knowledge about the phenomenon (Greene et al., 1989). Similarly, Punch (2005: 246) identifies three key points for consideration when

using mixed methods: whether the methods are taken as equal; whether or not they influence the operationalization of each other; and whether they are conducted simultaneously or sequentially.

Suggested benefits of mixed methods research include that it can increase the accuracy of research findings and the level of confidence in them generating new knowledge through a synthesis of the findings from approaches hearing different voices and bringing into play multiple constructions of the phenomenon (Moran-Ellis et al., 2006). Denscombe (2008: 272) suggests that this research methodology can increase the accuracy of data and provide a complete picture of the phenomenon under study than would be yielded by a single approach, thereby overcoming the weaknesses and biases of single approaches, and aiding sampling.

To summarise, the mixed methods research designs go beyond an exclusive affiliation to quantitative or qualitative methods and embrace a "pragmatist paradigm" (Teddlie and Tashakkori, 2009). These approaches draw on, and integrate both numeric and narrative approaches and data, where quantitative and qualitative methods are necessary and relevant, to meet the needs of the research rather than the allegiance or preference of the researcher, and in order to answer research questions fully (Johnson et al., 2007).

Creswell and Clark (2011), recommend that the researcher should make four key decisions in choosing an appropriate mixed methods design to use in a study, and these decisions are: (1) the level of interaction between the quantitative and qualitative strands;(2) the relative priority of the strands;(3) the timing of a strand;(4) the procedure for mixing strands. Furthermore, Creswell (2014) proposes three primary models for mixed methods designs: convergent parallel mixed methods, explanatory sequential mixed methods, and exploratory sequential mixed methods.

According to Creswell (2014: 15-16), in the convergent parallel mixed methods design, the researcher converges or merges quantitative and qualitative data in order to provide a comprehensive analysis of the research problem. In this design, the researcher collects both forms of data roughly at the same time, and then integrates the information in the

interpretation of the overall results. Contradictions are explained and discussed in this design. In the explanatory sequential mixed methods design, usually, the researcher first conducts quantitative research, analyses the results and then builds on the results, explaining them in more detail with qualitative research. It is popular in fields with a strong quantitative orientation. The exploratory sequential design is the reverse sequence from the explanatory sequential design. In this design, the researcher first begins a qualitative research phase and explores the views of participants. The data are then analysed and the information is used to develop the second, quantitative phase.

This study aims to investigate the conceptualisation, learning, use of formulaic language, and students' reflections on the teaching of formulaic language in an EFL context. It probes formulaic language learning and use by analysing and comparing both qualitative and quantitative data to gain an overall picture of the research site textual and perceptual data, rather than being either qualitative- or quantitative-orientated. Therefore, this research adopts convergent parallel mixed methods. Creswell (2009: 213) clarifies that in this model generally uses separate quantitative and qualitative methods as a means to offset the weaknesses inherent within one method with the strengths of the other. In this approach, the quantitative and qualitative data collection is concurrent, happening in one phase of the research study. Creswell also stresses that ideally, the weight is equal between the two methods, but often in practice, priority may be given to one or the other, thus it is a "side-by-side integration" (ibid.). The mixing of methods during this approach is usually found in an interpretation or discussion section.

Coming back to the research questions, the first set of research questions can be answered by analysing the textual data, which consists of the student written corpora by the corpus analytic tool (Wordsmith 6.0) (see section 5.8). In order to answer the first set of research questions, the researcher, on the one hand, employs an exploratory design to identify the formulaic language in students' texts first, i.e. through qualitative methods (i.e. students' perception-led identification), where the informant writers recognise the formulaic language that they used in their own written texts. On the other hand, the researcher uses the computer software to extract the pre-defined linguistic

clusters from the written corpora. As discussed in section 2.2, previous research on formulaic language has shown limitations in providing convincing methods of identifying formulaic language. With these mixed methods research designs, it is possible to expand the understanding of the formulaic language used by EFL learners by drawing on both the speaker-internal and -external approaches.

As for the second set of research questions, which target the student conceptualisations of formulaic language in English, they can be addressed by using both quantitative textual data (corpus analysis) and qualitative perceptual data (student recognition). Thus, at the stage of overall interpretation, the two strands of data are then compared and synthesised to provide a justified and well-rounded interpretation of the data. With regard to the third research question, this addresses the learning, use and teaching of formulaic language in English at the research site, which mainly draws on perceptual data collected from the semi-structured interviews.

Regarding criteria for the quality of mixed methods research, Dörnyei (2007:63) argues that in order to ensure a good design validity, researchers have to present evidence for two aspects: first, they have to justify the choice of the specific research methods that are combined in the study and present the main rationale with regard to ensuring a match of the research question or topic with the methods; this is summarised as the "fitness-for-purpose" selection. The second aspect is that the researchers have to demonstrate that the mixed design displays enhanced validity relative to constituent methods.

The research questions proposed in the last section (RQ 2 and RQ 3) aim to investigate the learners' language learning strategies and learners' perspectives on learning formulaic language. The researcher is going to use a triangulation design, which uses both quantitative methods (corpora analysis) and qualitative ones (the elicitation of students' recognition). Both types of data are collected in parallel and then synthesised to develop a more nuanced understanding formulaic language in EFL contexts.

5.2.2 Design of the study

As stated in Chapter 1, this study aims to explore and describe the learning and use of formulaic language by Chinese university students in their written texts, in order to further develop the understanding of formulaic language as a phenomenon in second language learning. More precisely, this research is going to examine formulaic language use in EFL learners' texts (RQ1a to 1d), the student conceptualisation of formulaic language in English (RQ2 and 2a), and the issues related to the learning, use and teaching of formulaic language (RQ3). A mixed methods design is employed in the study.

From the research questions, it can be seen that both strands (quantitative and qualitative research methods) are going to be employed in the research and they are going to be kept separate during design, data collection, and data analysis. In order to provide a justified and triangulated interpretation and conclusion, the results from both strands will be synthesised after the analysis of each method (see Chapters6, 7 and 8).

As stated in the research methods section (section 5.2.1), the exploratory approach is used to address the first research question; and the triangulation design is usually used to investigate the remaining two sets of research questions. In the research design, the qualitative approaches and quantitative approaches are intertwined in the data collection steps. The research design and data collection steps are shown in the figure below:

Data collection steps	**Notes**	**Date and time**
1) Early days on the research site: obtain informed consent of participants and establish rapport with students; gain background information about participants	When entering the research site on March 8, 2013, I aimed to obtain more information, such as the environment, students, and teaching staff. I was already familiar with the context because I studied for my BA degree there for four years and grew up in that city.	Weeks 1 and 2
2) Select the 12 focused participants and collect the first written texts	From information gained through Step 1), the researcher chose the 12 students as the informant writers for the follow-up interviews and the identification of formulaic strings. The researcher collected the first writing tasks from all the participants and picked up the 12	Week 3

	students' written texts before teachers marked them.	
3) Conduct first round of interviews	Shortly (no more than two days in general) after the writing tasks were finished by the students, the first interviews were conducted. During these interviews, the 12 informant writers were then asked to mark the formulaic language in their written texts before the interview started or at the end of the interview. The questions in the interviews are about general language learning experience regarding formulaic language (see Appendices 3 and 4).	Week 6
4) Collect the second round of written texts with marked formulaic language	The 12 informant writers marked the formulaic strings as soon as finished writing.	Week 9
5) Collect the third round of written texts with marked formulaic language	The 12 informant writers marked the formulaic strings as soon as finished writing.	Week 12
6) Conduct the second interviews	The second round of interviews were conducted and were designed to examine the specific formulaic language learning strategies used in the composition of English texts (see Appendices 5 and 6), as well as the students' viewpoints on formulaic language. Interviews were conducted shortly after the exam answer sheets (see Figures 5.6 and 5.7) were collected	Week 16

Figure 5.2: Data collection steps

All four sets of written texts from two groups of student writers were used to build up the corpora for further examination of the linguistic structures and discourse functions of formulaic language used in EFL learners' texts in detail.

In these six steps, both quantitative and qualitative methods were used to examine the learning and use of formulaic language in student writers' language. Two types of data were collected, i.e. the elicited perceptual data and naturalistic textual.

As stated above, the design of this study synthesises qualitative and quantitative methods. Through the thick and rich description of the participants' perspectives from the qualitative approach, and numerical information collected by the quantitative approach, the picture of formulaic language learning and use by Chinese EFL learners can be portrayed.

5.3 Participants

There are two main groups of sampling strategies: probability sampling and non-probability sampling. Probability sampling is usually beyond Applied Linguistics, and is mostly used in scientific procedures, including random sampling, stratified random sampling and systematic sampling, and cluster sampling, while non-probability sampling is purposive, usually conducted in three ways, namely, quota sampling and dimensional sampling, snowball sampling, and convenience or opportunity sampling (Dörnyei, 2007). In this research, the participants consist of students from Year 1 and Year 3 in the foreign language department. They were selected as the participants mainly on the basis of convenience or opportunity for sampling strategies, which ensured that members of the target population were selected for the purpose of the study can meet the practical criteria, i.e. geographical proximity, availability at a certain time, easy accessibility, or willingness to volunteer.

Regarding the student participants, as mentioned in the local context of the university (section 4.3.2), there are about 400 enrolled students in the department of foreign languages. Based on the research questions (1d and 2a), the study is going to make comparisons between learners at different proficiency levels, so participants enrolled in different school years were chosen as they are expected to be with different proficiency levels in the English language. Therefore, I decided to choose a group from the senior year (Year 3 or 4) and the junior year (Year 1 or 2) respectively.

It was the second semester of the academic year (2012-2013), and TEM-4 held in April and TEM-8 in March for Year 2 and Year 4 students respectively. Normally, the department stops the classes or lectures until the tests are finished, in order to provide more flexible time for students to prepare for the English major tests. Thus, the class organisation changes and it is not easy to reach the students from these two school years. Besides the test, the final year students graduate in summer, there are no credited courses on their timetable, and the university encourages them to find a job or internship as soon as possible. Some students even start working before they graduate.

Compared to students in their second and final years, the Year 1 students are more willing to take part in English language activities, for example, the English society on campus. They are also more willing to participate in some academic research with the lecturers as a newcomer to the academic community. With enthusiasm and curiosity about the university life and research in the field of their study, they were more willing to join in this research study. Similar to this, Year 3 students agreed to be the participants for the research.

Since the corpora needed to be compiled to investigate formulaic language use in learners' written texts, it was necessary to collect sufficient written texts on the same topics. Therefore, the students enrolled in the course *A Basic Course in English Writing* (Year 1 students) and *Advanced English* (Year 3 students) were invited to take part in the research.

Initially, the research was designed to choose 12 informant writers to have follow-up interviews with equality in the gender of the participants. However, in the foreign language department, there are fewer males than females, and the males are much less motivated than the females. Thus it is difficult to maintain the data collection for a relatively long time, even if they agreed to participate in the first interview. In order to maintain the long-term data collection, according to time availability and willingness, four male and eight female students were selected to be the informants.

5.4 Researcher's role

The researcher plays a pivotal role in the research process, especially in qualitative research methods, as data are collected, information gathered, settings viewed, and realities constructed through the researcher. Further, the researcher is responsible for analysing the data through an iterative process that moves back and forth between data collected and data analysed (Lichtman, 2012: 21). Finally, the qualitative researcher interprets and makes sense of the data (Coffey and Atkinson, 1996), whereas the quantitative researcher is more likely to study samples or the subjects, which are usually nameless and faceless individuals who have been chosen at random to represent others

with similar characteristics (Lichtman, 2012). Thus, the role of the quantitative researcher in the actual analysis is relatively limited.

As this study employs a mixed methods approach, on one hand, the researcher deals with relatively large samples, for example, students' written corpora, and on the other hand, the researcher carries out some qualitative surveys, i.e. interviews, at the research site. Especially during the fieldwork, the researcher's role is eminently important, since the success of data collection during fieldwork depends on whether the researcher is able to define and negotiate his/her position with all participants. Therefore, it seems necessary to clarify the researcher's role in this type of research.

Saragni and Candlin (2003) discuss the different roles that the researcher might adopt in applied research, that is, researcher as outsider/insider; researcher as resource; researcher as befriender; researcher as target audience and assessor of performance; researcher as expert/consultant and agent of change. During the fieldwork, the researcher enacted multiple roles in the setting, sometimes concurrently. Thus, the researcher approached participants and helped them with language learning, she acted as source of knowledge about universities and English language. As a former student of the department, the researcher was invited to give some talks, meet with students and share her experience, which decreased the distance between the researcher and the students, this leads her role was as befriender. The researcher also acted as assessor of student performance, as a teacher, on certain modules across classes and disciplines, such as *Intensive Writing Training Course for TEM-4* and the *English Movie Appreciation* and *Preparatory Course for College English Tests*, and also as a supervisor and examiner for some undergraduate dissertations.

The researcher did not teach the participating students, in this way, the researcher enables keep her role as "outsider". At the same time, by taking on those roles, it was conducive to establishing links with the staff and students in the setting. This was in addition to the researcher's general familiarity with the setting based on her previous experience as a BA student in the department.

All of the above helped the researcher to maintain effective relations with participants as well as partially develop an “insider” perspective. Following Richards (2003), this positioning can be described as “emic” and/or “etic”, which are terms sometimes used to define an insider’s perspective on events (emic) as opposed to an outsider’s one (etic) and it is important to maintain a balance between the two. In order to achieve this balance of different perspectives on the events during fieldwork, the researcher asked fellow students in the UK and staff at the research site to assist with the trial of the interviews. The researcher also maintained contact with her supervisor via email, reporting the data collection progress.

5.5 Fieldwork

The fieldwork spanned a total period of 5 months (one semester), starting in late February 2013 and ending in early July 2013. On arrival, the researcher’s first step was to obtain some general information about the participating students and to establish a relationship with staff in the department. In mid-March, i.e. 2 weeks into the research phase, the researcher sent out the participant information and consent forms to the Year 1 and Year 3 students. On the basis of the feedback from the consent forms and teachers’ suggestions, 12 students were selected as the informants for the interviews.

The first interview was conducted one week after the students finished their first assignment, i.e. in week 3 of the research. Over the next few weeks, i.e. April and May, 2013, the researcher audited some classes in order to collect the students’ written text, which they handed in as homework in the class. During that period, informal discussions with the teachers about the arrangements for the homework assignments took place. In June, the second interviews were conducted after the final exams of that semester.

During these five months, the researcher developed a deeper understanding of the students and the staff at this university. At the same time, the researcher also managed to establish a very friendly relationship with the staff team and students at the research site. This decreased the distance between the researcher and the participants, which had beneficial effects on the quality of the interviews conducted.

5.6 Research instruments

In this study, two main research instruments are used to collect the data, namely, students' written texts and two rounds of semi-structured interviews.

5.6.1 Students' written texts

When talking about issues in writing research (see Chapter 3), Hyland (2009) notes that writing research tends to favour data gathered in naturalistic rather than controlled conditions. While methods that elicit data through structured interviews or experiments can provide interesting and useful insights into writing, data collected via observation or authentic texts are more common.

Hyland (2009) argues that the major source of data for writing research is writing itself. Thus, in order to answer RQ 1, the corpora of students' written texts had to be built up to facilitate the analysis of student formulaic language use in texts.

As discussed in section 3.2.2, the majority current research on formulaic language in Chinese university students' writing by the corpus analytic approach either is based on large corpora data without distinguishing the proficiency levels of the writer (Chen, 2009; Chen and Baker, 2010); or focuses on general population of non-native speakers with different L1s (e.g. Biber, 2006; Biber, Conrad and Cortes, 2004; Mauranen, 2012). As for the Chinese writer-only corpora, their access and publications are very limited. Moreover, these studies collected texts from writers at a single point of time rather than over a period of time. Therefore, it is reasonable to build up the written corpora by collecting written texts at the research site over a period of time. According to the design for this research (see section 5.2.2), this study will synthesise textual and perceptual data to explore formulaic language use in Chinese university students' writing, with a focus on the differences across proficiency levels (RQ 1a to 1d).

Due to the lack of English practices at the research site (see Chapter 4), in order to ensure that sufficient written texts can be collected and the students received the

similar writing tasks and submitted their texts at the same time, the researcher chose two compulsory modules from Year 1 and Year 3 respectively.

5.6.1.1 Year 1 students' written texts

Overall, 83 Year 1 students who were registered for the module *A Basic Course in English Writing* were chosen for the study. This module emphasised English writing and students had intensive practice on producing English texts. This module ran 16 weeks for one semester, with three writing tasks as homework and one written task in the semester final exam.

Turning to the texts produced as homework, these writing tasks were given to students at the end of a class or on a slide/projection. Students were allowed to use reference material, including dictionaries, books, the internet, and or seek for support from native speakers. The students had to submit their homework assignments at the beginning of the next lecture. Samples of the Year 1 student written texts for each task can be found in Appendix 7.

The first homework was a 100-word essay on "self-introduction". The teacher required students to write something about themselves, such as their experiences, hobbies, and dreams. Through this task, the lecturer could start to get to know the students in the class.

The second task asked students to write a 100-word text about their own experiences. The covert aim of this task was to check student attendance after a 4-day public holiday (Qing Ming Festival). However, the teacher did not provide a clear explanation of the specific scope of the experience. Thus, it is not surprising that in the collected texts, some students just wrote some stories about their childhood, some wrote about the things they had done in the past few days, others recounted their whole adolescence, including which schools they had attended and how they were admitted to this university.

The third task was a short writing text on “the benefits of volunteering”. The students were expected to write about 200 words. The lecturer said that this task would give students a taste of the writing task in TEM 4. The prompt for this writing task was projected and read as follows:

Direction：

December 5th is International Volunteers Day. Since 1985, when the United Nations announced the special day, tens of millions of people around the world have volunteered to help those in need. China now has 4.5 million registered volunteers how have provided more than 4.5 billion hours of volunteer work. What can you gain from volunteering?

Write a composition of about 200 words on the following topic:

The Benefits of Volunteering

You are to write in three parts:

In the first part, state clearly what your view is.

In the second part, support your view with appropriate reasons.

In the last part, bring what you have written to a natural conclusion or a summary.

Marks will be awarded for content, categorisation, grammar and appropriateness. Failure to follow the instructions may result in a loss of marks

Figure 5.3: Prompt for writing task: the benefits of volunteering

The exam constituted the fourth task. The prompt and task requirements are shown in Figure 5.4 below:

Part two
The Dragon Boat Festival is one of the most important traditional festivals in China. Write a composition about 200 words on the following topic:

The Dragon Boat Festival
First, you should say what you know about the festival.
Second, you should describe how you or other people usually observe the festival
In the last part, bring what you have written to a natural conclusion or a summary.

Marks will be awarded content, categorisation, language and appropriateness. Failure to follow the instructions may result in a loss of mark.

Figure 5.4: Prompt for writing task: the dragon boat festival

The prompt for this composition is taken from a previous TEM-4 exam. This task is in the second part of the semester in the final exam of the writing course. In the semester final exams, students were much more anxious, as they had to finish writing in time together with other tasks in the exam papers, without access to any reference material, including dictionaries, textbook, or the internet. Also, this writing contributes to their course assessment largely, therefore, the students were generally very motivated.

The table below shows the overview written texts collected by the researcher from the Year 1 students.

Table 5.1: Written texts collected from Year 1 students

Topic	**Types**	**Aims**	**Words Length No.**	**Numbers of texts**
Self-introduction	homework	Getting to know students	100	83
My experience	homework	A taste of TEM-4 test	200	73
The benefits of volunteering	homework	Attendance check	100	61
The dragon Boat Festival	exams	TEM-4 test preparation	200	83
Total Number of words in text corpus			56058	300

5.6.1.2 Year 3 students' written texts

Overall, the texts written by 73 students in the third year signed up for the module of *Advanced English* were chosen for corpora compilation (see section 5.7). The module emphasises the improvement of English language skills for students in the senior years at the university. Similar to Year 1 students, the writing topics are usually given at the end of the lecture and the requirements explained in Chinese. Teachers stressed the importance of the structure and language of the texts, as well as the word limit, which was usually about 250 to 300 words. Samples of the Year 3 student written texts for each task can be found in Appendix 7.

The first writing topic used was 'Should we help strangers?', and students were required to write 300 words on this as homework. This task aimed to discuss a phenomenon in current society and followed on from a lecture about helping others.

The second task asked students to write a 200-word text about a Chinese traditional festival, Qing Ming (Tomb Sweeping). This public holiday takes place on April 5th each year, when people go outside and enjoy the greenery of springtime and go to graves to remember departed family members. This writing task was taken from a previous TEM-8 mock test.

The third writing task constituted a 200-word book or a movie review. The covert aim again was to check student attendance after a public holiday (Qing Ming Festival), which resulted in somewhat vague task requirements where the teacher did not set a specific book or movie.

The fourth writing task was a part of the final exam of the semester for the course *Advanced English*. The prompt for this writing task is shown in Figure 5.5 below:

Writing
Directions: For this part, you are allowed 30 minutes to write a short essay entitled **The Importance of Reading Classics**. You should write at least 300 words following the outline given below.

阅读经典书籍对人的成长至关重要
现在愿意阅读经典的人越来越少，原因是……
我们大学生应该怎么做

Figure 5.5: Prompt for writing task: the importance of reading classics

The Chinese in the paper is the actual writing prompt, which translates as

1. The reading of classics is important for one's life;
2. The number of people who are willing to read the classics is decreasing, the reasons are…
3. What can we do as university students?

During the exam, the students were asked to finish the writing and other tasks on the exam paper, without resorting to reference materials or other resources. This task was

chosen for the writing task in a past College English Test 6 (CET 6). The teacher who designed this exam paper thought it was too difficult for the students to finish a TEM-8 writing task at that time, and so chose the easier task from the CET6.

An overview of all the texts that the researcher collected from the Year 3 students at the research site is provided in the tables below:

Table 5.2: Written texts collected from Year 3 students

Topic	**Types**	**Aims**	**Words Length No.**	**Numbers of texts**
Should we help strangers?	homework	TEM-8 test preparation	300	65
Qing Ming Festival	homework	Mock test	200	54
A book/movie review	homework	Attendance check	150	73
The importance of reading classics	exams	TEM-8 test preparation	300	59
Total Number of words in text corpus			64186	251

In addition to collecting student texts, the 12 informant writers marked what they considered as formulaic language in their texts once finished writing. As for the exam writing, the participants marked formulaic language on the photocopies of their exam paper once they finished the exams. In this way, it could be ensured that these informants had as accurate an impression of their use of formulaic language as possible.

All the written texts collected in this study are naturalistic, as they were completed by the students for external purposes rather than for the compilation of the corpora. These texts are argumentative or narrative essays. The topics of these written texts cover a range of accessible, real-world knowledge. It is noted that all the written texts collected for this study are relatively short. There is no textual plurality presented in these texts (Moore and Morton, 2005), in other words, these tasks do not require background reading or research, nor do the texts cite or refer to the work of others. Rather, the texts consist of prose-only responses to set titles soliciting the writer's opinions on a general knowledge topic or personal experience (Leedham, 2011; Moore and Morton, 2005;

Leedham, 2011). In theory, each student writer was supposed to contribute four texts. However, in practice, students can be absent from the class or change the module after first few weeks. Therefore, each student may actually contribute between zero to four texts.

As for the time for writing, little preparation time was given when writing was conducted in class or exams; whereas writings for homework were set over an extended time period within one or two weeks. During this period of time, students could take as much time as they wish on preparation, drafting or consulting to any references, or seeking for support from the native speakers. Students handwrote their texts. The researcher kept the original copies of the homework tasks and photocopies of the exam tasks. These written texts were then keyed in for the purpose of corpus compilation.

Students were highly motivated if they were producing writing for assessments which contribute to their degree, for example, for exams or mock-test purposes; whereas their motivation was unclear when the texts were written in class but were not assessed, for instance, for attendance checking.

5.6.2 Interviews

Traditionally, interviews are categorised in terms of the question structures, participant organisation, and the formality of conducting the interviews. Based on the way of organising participants, interviews can be conducted as individual or group interviews (for example, focus groups); with regard to the degree of formality, there are formal and informal interviews. The most popular way to categorize interviews is according to the structures of questions in the interview, grouped into four types: 1) informal conversational interview; 2) interview guide approach; 3) standardized open-ended interview 4) closed, fixed response interview (Patton, 1980: 206). However, Nunan (1992) claims that most interviews can be placed on a continuum ranging from unstructured to structured, with semi-structured interviews taking the middle ground.

Structured interviews, or standardised interviews, can be seen as a spoken questionnaire (Dörnyei, 2007, Talmy and Richard, 2011) since the researcher follows a

pre-prepared, elaborate interview schedule/ guide which contains a list of questions to be covered closely with every interviewee. This type of interview covers a well-defined domain, focused target topic, as well as providing comparable answers across different respondents. However, there is little room for variation in the responses. At the other extreme, the unstructured or "open interview", also referred to as an "ethnographic interview", allows maximum flexibility to follow the interviewee in unpredictable directions, with only minimal interference from any research schedule (Dörnyei, 2007). A drawback of this kind of interview, as Talmy and Richard (2011) argue, is that there is no fully "open" interview in the sense that the interviewer goes in and simply encourages the respondent to start speaking, ready for anything that may emerge. They further claim that respondents who do not have a clear idea of the researcher's interest are likely to expend a lot of energy trying to figure these out, and the researchers are usually able to recognise and draw out the emergent structures.

Semi-structured interviews, as indicated in their name, can be seen as a mid-point between the two extremes. There is a set of pre-prepared guiding questions and prompts, the format is open-ended and the interviewee is encouraged to elaborate on related issues in an exploratory way. As for the advantages of this research instrument, Dörnyei (2007) concludes that it is a natural, comfortable, and socially acceptable way to get information and can be used in various situations as well as on different topics to elicit in-depth data; and the interviewer can use flexible approaches to probe into issues with the systematic interview guide. Mackey and Gass (2005) mention that it allows researchers to investigate phenomena that are not directly observable, in the case of learners' self- reported perceptions or attitudes. Also, there are some drawbacks to this research instrument. For example, besides requiring good communication skills from researchers, this interview format does not allow for anonymity, and so the respondent may display him/herself as 'better' than in reality. This could result in the "halo effect" which refers to what happens when interviewees pick up cues from the researcher related to what they think the researcher wants them to say, thus potentially influencing their responses (Mackey and Gass, 2005), or bias caused by the asymmetrical relationship between interviewer and interviewee (Nunan, 1992).

Parts of this research project, such as RQ 3, are clearly exploratory and qualitative, with the aim to describe the perception of formulaic language and its role in language learning. Therefore, semi-structured interviews were used to gain in-depth information from the interviewees. The interviewees were chosen in this study was mainly according to their availability and willingness to take part in two rounds of interviews over a period of 16 weeks (see sections 5.2.2 and 5.3).

Polkinghorne (2005) argues that one-shot interviews are rarely able to produce the full and rich descriptions necessary for worthwhile findings. In order to address this issue, the design of two rounds of interviews in this study ensures the validity and reliability of the research methods and data collection (Maxwell, 1992).

Both interviews were semi-structured, allowing for certain questions to be changed in order, but the informants were asked the same questions in general (Dörnyei, 2007, Talmy and Richard, 2011). The researcher followed Dörnyei's (2007) suggestion that the whole interview schedule should consist of four kinds of questions, that is, opening questions, content questions, and probe and ending question. Opening questions refer to the first few questions, usually used to set the tone and create an initial rapport, such as some easy personal or factual questions. For example, in both schedules, the participants' personal information (e.g. name, class) was always checked first. Content questions usually concern different aspects of the participants' overall view or experience of the topic and therefore the researcher can get a rounded picture. In this part, the questions are mainly about experience and behaviour, opinions and values, feelings, knowledge, or background information and demographic information. With regard to the probe questions, the researcher can use what the interviewee has said as a starting point to go further and to increase the richness and depth of the responses, which usually takes the form of detail-oriented or clarification questions. For example, in the first interview schedule, the question "How do you deal with the highly repeated formulaic language in writing?" was followed by "Why questions?" as a typical probe. This combines a detail-oriented and a clarification question. The ending questions function as a signpost to show the interviewees that the interview is nearing completion.

The first interview outlined the scope of the research for the interviewees (see Appendices 3 and 4). The questions ranged from broad ones on English language learning to detailed ones on formulaic language. In this way, the interviewees became aware of what type of questions the researcher was going to ask, which allowed them to recall some information on the topic. In the second interview, however, more detailed information on the learning and use of formulaic language was elicited. In the first interview schedule, taking into account that it was not long after the researcher had arrived at the research site, it was possible that the researcher would not be yet familiar with the participants. To decrease the distance in communication between researcher and informants, the ice-breaking phase in the first interview is very important. Therefore, there are more warm-up questions in the opening than in the following interviews. In total, there are 17 questions in the first interview. Among them, 11 are content questions and 4 are probes on formulaic language. The interview took an average of 25 minutes each.

After one semester, the researcher and participants had got to know each other better, and so it was possible to ask some more detailed questions of the participants. At the end of the semester, the second interview took place, containing 11 questions. Four of them are content questions and five probes. It took about 13 minutes to complete each interview on average. English and Chinese versions of the schedules for both interviews are shown in Appendices 3 to 6. In this way, the research administered two rounds of interviews with the same interviewees to obtain sufficient depth and breath. As some questions in the second round of interviews were further developed from the first round, with a narrower focus on the core topics of the research. It enables a “deeper digging” to minimise the interviewees’ tiredness with the questions (Stebbins, 2001).

While conducting these interviews, the researcher provided both English and Chinese versions of the schedules, and the informants made the decision as to whether the interview was conducted in English or in their first language, Mandarin Chinese, according to which language made them feel more comfortable. On average, two rounds interviews seemed relatively short, however, the use of the shared L1 shorten the distance between the researcher and the interviewees and information delivered

with high efficiency. Besides this, the interviews took place individually without other teachers being present. There were 12 informants for each round of interviews, and the researcher conducted 24 interviews in total with the same 12 interviewees. All the interviews were audio-recorded. General information about the interviews is illustrated in Table 5.3. All participants have been given pseudonyms in the research.

Table 5.3: Information on interviews

Interviewees	**1st interview time length**	**2nd interview time length**	**Interviewees**	**1st interview time length**	**2nd interview time length**
INV001	27	11	INV007	19	16
INV002	26	19	INV008	25	6
INV003	14	7	INV009	20	11
INV004	30	24	INV010	44	24
INV005	30	15	INV011	17	10
INV006	30	13	INV012	15	11
Total time	157	76	**Total time**	140	79

(Time: min)

5.7 Data types and overview

Two types of data were collected for the study: student written texts and interview data. As the primary data collected for this study, students' written texts were collected over one semester (16 weeks) to assemble the corpus, which consists of two sub-corpora: the Year 1 and Year 3. The tables below show an overview of the student written corpora. It is noted that these are two very small size corpora; the student writers are from a very homogeneous language background; texts produced by the students are generally short, only 187 words in a Year 1 text and 255 words in a Year 3 text on average.

Table 5.4: Overview of student written corpora

	No. of students	**Copies of texts**	**No. of words**	**Average Length of a text**
Year 1	83	300	56058	187
Year 3	73	251	64186	256
total No.	**156**	**551**	**120244**	**221.5**

The researcher then keyed all the handwritten texts into electronic MS word documents in order to enable the use of corpus analytic software (WordSmith 6.0). During this

process, incorrect spellings were edited. However, only two specific cases in the written text samples were considered as misspellings to be corrected. These were: 1) a mixture of American and British English spelling, like *colour* (*color*); 2) wrong spellings, for example, *silent* (*scilent*). In order to keep the texts as close to the original as possible, all other types of grammatical errors or mistakes were kept during the transfer process, such as article an/a, singular/ plural form, and subject-object agreement issues.

As for the interviews, in total, the two rounds of interviews were audio recorded with 12 interviewees. The total length of all recording was 7 hours and 44 minutes.

Table 5.5: Overview of interview data

		Length of time (approx.)		**Average time**	
	No. of students	**1st interview**	**2nd interview**	**1st interview**	**2nd interview**
Year 1	6	157min	76min	26min	13min
Year 3	6	140min	79min	23min	13min
Total No.	**12**	**297min**	**155min**	**25min**	**13min**

Both rounds of interviews were conducted with the same 12 interviewees (six from Year 1 and six from Year 3), and the related background information is shown in the table as follows:

Table 5.6: The overview of the interviews

School Year	**Name**	**Sex**	**Hometown**	**Age**
Year 1	INV 001	M	Hubei Jinmen	18
	INV 002	M	Hubei Xiantao	21
	INV 003	F	Hubei Wuhan	18
	INV 004	M	Hubei Shiyan	19
	INV 005	F	Henan Xinyang	19
	INV 006	F	Hubei Shiyan	19
Year 3	INV 007	M	Hubei Shiyan	21
	INV 008	F	Hubei Shiyan	24
	INV 009	F	Hubei Shiyan	23
	INV 010	F	Hubei Shiyan	22
	INV 011	F	Hubei Shiyan	22
	INV 012	F	Henan Xinyang	22

The table above shows that the 12 interviewees were from the local area, Hubei Province, or the province, Henan, which is near the one in which the university is situated. The dialects in these areas are very similar. Thus the 12 interviewees lived in areas with similar dialectal, economic and educational conditions. Also, the rounds of interviews in this research recruited the interviewees from both genders. As for the age of the interviewees, this ranged from 19 to 24 years old. For the interviewees from the Year 1 group, their ages were between 19 to 21 years old; and in the Year 3 group, the interviewees aged from 21 to 24 years old. The reasons for the different ages in the same school year group was that the interviewees entered the school at different ages. Some of them started school earlier, some of them later. In addition, none of them had worked before they entered university; in other words, they had been classroom learners of English language all the time. Thus, it is presumed that all 12 interviewees had received a similar education in the English language, as well as being in a similar age group.

By this practice, it can ensure that within this single linguistic community, the participants in this study shared very similar "verbal repertoire" (see section 2.2.2.2), including dialectal and register factors, as well as similar social situation. These are very important factors to be taken into account when research formulaic language in EFL contexts from psycholinguistic and cognitive perspectives. The guide questions for the interviews can be seen in Appendices 3 to 6.

5.8 Data analysis

As mentioned in the previous section on the research design, this study employs a mixed methods design to investigate formulaic language use, learning and teaching at the research site, as well as the student conceptualisation of formulaic language. The main data are students' written texts and semi-structured interviews with the 12 student writers. Therefore, data analysis consists of two parts: the first part focuses on written text analysis; and the second part on interview analysis. Next, the analytic procedures for each part will be introduced in turn.

5.8.1 Written texts analysis procedures

The methods of analysis used in the written text analysis are firmly grounded in a combination of a students' perception-led approach and a corpus-based approach. The analytic procedure for students' written texts involves three phases.

The First Phase constitutes the analysis of the formulaic strings identified by the students according to their perception. Phase 2 is the analysis of four-word strings based on the frequency count in student written corpora. The first two phases are mainly quantitative. Using these methods allows for a detailed comparison of different identification approaches based on quantitative evidence and also permits more in-depth qualitative analysis of certain items chosen, based on those quantitative findings, in Phase 3. It is noted that this study aims to describe the formulaic language used in the students' texts by two independent identification approaches, therefore, this study focuses on the types (different strings/clusters) rather than the tokens (total occurrences of strings/clusters) (for more discussion on types and tokens see Chen and Baker, 2010; Huang, 2015).

Phase 3 involves the comparison of the results generated from Phase 1 and Phase 2. At the same time, the researcher comments on the methodological issues regarding the identified formulaic language. This approach is thus suited for the analysis of formulaic language, both from cross-corpus and corpus internal comparative perspectives. Ultimately, the most revealing insights into formulaic language use will be gained from a closer look at the texts, the speakers, and the situational variables. As Simpson (2004:41) argues, quantitative analysis alone can never provide a satisfactory picture, especially when one of the research aims is to make the findings valuable to language teaching.

Phase 1: Student identified formulaic expressions analysis

Twelve interviewees were asked to mark what they considered to constitute formulaic language in their own English written texts as soon as finished writing. The working definition of formulaic strings, i.e. multi-word psycholinguistic units, was given to the students to help to better understand what constitutes formulaic language. The emphasis was put on the these commonly used strings were learned and memorised as whole rather than constructed word by word, regardless of language native norms. All

the formulaic sequences identified were regarded as legitimate in the sense of reflecting student perceptions. Figures 5.6 and 5.7 below illustrate two samples that show how students marked formulaic sequences in their texts.

As it's known to us all, December 5th is International Volunteers Day. **To be a** volunteer **is** really a good and meaningful thing in that you can gain a lot from it, such as **making new friends with** others, putting **what you have learnt** into practice, **to improve your skills and so forth**.

When you are going to be a volunteer, you will do **something helpful** to others, so maybe you will make friends with **those who** you help and those who are volunteers just like you. When you go to the west of China to support that local education, you'll **put** what you have learnt in class **into practice**. **I think it's a** really exciting and interesting thing to use your knowledge to help rural area children. **What's more**, if you are a volunteer, **you will** improve your skills. **For instance**, thinking you are a Olympic Games volunteer, **you are due to** meet many people. **Some of** them will communicate with you, others may need your help. So **you will** improve your skills of communicating, ability to do things.

In a word, **to be a volunteer** is a really beneficial thing. **If you** join the volunteer team, **you will** surprisingly find you are not what you **used to be, in other words**, you will be a kinder, experienced, knowledgeable person.

Figure 5.6: A written text sample from a Year 1 student

Recently, **there is a hot issue** that is widely debated, the topic is: whether should we help strangers? **In terms of** this, I have something to say.

Indeed, these cases which the strangers cheat or hurt the one who have helped them before exist, and **due to** the widespread of media, these case are being magnified; thus causing a worldwide concern on themselves' safety but those phenomena are **in some degree** on the last part. **In my view**, I still think that we should help strangers, **here is as follows**.

For thousands of hundreds years, we have been taught acting nice to everyone. Thus, we win such a harmony (hamony) society, the development of society need a warm atmosphere too. If we concentrate too much on the bad aspect, we won't worry ourselves all the time, we thus cannot **spare our time enjoying** the beauty around us. It's not the true meaning of life at all, so going our hands to these who need help. It is not only helping other, it helps ourselves too, the energy of warm can be transferred through our hands. Those repay their helpers badly can **sooner or later** shamed of what they did, apart from this, when we help strangers, we

should **more or less** be aware (ware) of getting cheated and use common sense to tell them, after doing this, we definitely (definately) can gain ourselves satisfactory!

Figure 5.7: A written text sample from a Year 3 student

As can be seen in the figures above, the expressions in Bold were considered formulaic by the informant writers. These sequences were marked as soon as these informants finished the writing.

These formulaic expressions include continuous and discontinuous strings and the distinction between continuous and discontinuous strings is mainly according to the use of ellipse marks by the students. For example, the string *only when…can we* is discontinuous since there is an ellipse mark include, in other words, some open slots taken up by space in the strings.

The length of those identified strings varies from one word (individual words) to more than seven words. In general, more than seven words strings are usually whole sentences, such as formula like, **only the deaf appreciate hearing, only the blind realise the manifold blessing that life in sight (only the deaf appreciate hearing, only the blind realise the manifold blessings that lie in sight).* The one-word expressions are mostly conjunction words, such as *but, nevertheless, besides, though, however;* or hyphenated compounds like *short-time, ex-welder, grief-stricken.* It is also noted that in the student perceptions, formulaic language can range from a micro-level, for example, a single word (e.g. *however, except, ex-welder*) to a macro-level, for instance, the schematic structure of a paragraph or a whole written text (*Firstly…Secondly…Thirdly…*) (see Appendices 7 and 10).

Some strings are content- and topic related, which often were directly copied from the writing prompts or related to the specific background of the topic. For example, there was a task that asked students to write a text about the *Qing Ming Festival*. So strings like *Pure Brightness Festival, Tomb Sweeping Day* (alternative translation for *Qing Ming Festival*) are considered as content-dependent strings. In the figures above, strings like, *to be a volunteer* and *spare our time enjoying* are considered content-dependent strings, therefore they are not included in the lists for analysis. It is also worth mentioning that,

as stated in the research instruments (section 5.6.1), Qing Ming Festival is a traditional festival and an important day of sacrifice for most people to commemorate their ancestors. The popular activities on this day include tomb sweeping, spring outing, and kite flying. Also, people usually have cold food instead of hot meals on this day. Some identified strings, therefore, contained information relating to this background, and examples include **set fire for food*, *ancestral tablets*, and *the tradition has been endowed with*.

It is clear that most identified formulaic sequences are multi-word (at least two words) semantic / functional units, regardless of nativeness or grammaticality. For example, the expression**for thousands of hundreds years* (*for hundreds of thousands of years*), is produced by the student and considered as a formulaic sequence, according to this student's judgement.

There are some strings that are obviously ungrammatical, however, still identified by the informant writers as formulaic strings, instances such as **rose to lover, smell in hands*, **waste- time and waste-energy thing* and ***for thousands of hundreds years*. The students recognised the first two instances (**rose to lover, smell in hands* and **waste-time and waste-energy thing*) as formulaic language probably because these strings have an equivalent formulaic expression in Chinese: *zeng ren mei gui, shou you yu xiang* (赠人玫瑰，手有余香) and *wang fei gong fu* (枉费工夫). The correct form for the last string (**for thousands of hundreds years*) should be **for hundreds of thousands of years*. The error occurred presumably because the student wanted to express the meaning in Chinese *qian bai nian lai* (千百年来), in which the word *hundreds* (*bai*) comes before *thousands* (*qian*). It seems apparent in these examples that students' English formulaic language use is strongly influenced by the conventionalised L1 use, and some assume a direct transferability.

In order to conduct a thorough and comprehensive analysis of the student-identified strings, I include all the strings marked by the students except those in the following instances:

First, individual words (mainly consist of adverbial linking words and hyphenated compounds) that are less than two words in length, given the focus of this research focuses on multi-words units .

Second, repeated identified strings or overlapping strings were also excluded from the analysis, as this study investigates the types (different types) rather than the tokens (total numbers) of formulaic language. For example, repeated strings like *rather than* were identified as formulaic strings twice by the Year 3 students, so there is only one instance of *rather than* as a formulaic string in the list. Overlapping strings like *I hope* and *I hope I'll*, where the former subsumes the latter, are considered as one type of formulaic string, *I hope I'll*. The tentative conclusion can be drawn that those subsumed formulaic strings are mostly caused by the individual language user's idiosyncrasies, for instance, the expression, **having delivered myself about* is taken as a subsumed string to "**having delivered myself about the*", which is ungrammatical and unlikely to be used widely within the discourse community in this study.

For the Year 1 group, after excluding the individual words and hyphenated compounds (11 types), and overlapping strings (17 types, including 5 types of subsumed types), the number of Year 1 student-identified strings in the final list is 210; whereas for Year 3, after the exclusions outlined above (11 types of individual words and hyphenated compounds; 18 cases of overlapped strings, including 4 types of subsumed types), there are 220 student-identified strings. The excluded items, i.e. the cases of individual words, overlapping and subsumed strings are unnumbered in the list. The refined lists of student-identified strings are illustrated in Appendix 10.

Phase 2: Four- word clusters analysis

The second phase is the analysis of lexical clusters based on a frequency-driven approach. The multi-word units of analysis for this study are frequently occurring expressions of four words, which I refer to as four-word clusters. As the clusters are identified by using a frequency-driven approach, the first step is to choose the cut-off frequency, which determines the number of bundles in the list for analysis. In previous research on clusters, n-grams and lexical bundles, the cut-off frequency has ranged from 10 times (Biber, et al., 1999; Biber, 2006), to 20 times (Cortes, 2004; Hyland, 2008), to 40

times (Biber, Conrad, and Cortes, 2004) per million words in the large size corpora, or has been converted to raw frequency according to a normalised rate, for example, in Hüttner's (2008) study, which sets 3 occurrences is set as the minimal raw frequency. The two sub-corpora investigated in the current study are of relatively small sizes with slightly different total word numbers. For the sake of comparison, this study applies raw frequencies converted from a normalised rate. So the raw frequencies for the two sub-corpora are 2 and 3 occurrences for the Year 1 and Year 3 corpora, as shown in the table below:

Table 5.7: Raw frequencies converted from a normalised rate

	Year 1	**Year 3**
Cut-off point in absolute frequency	2 times in 56058 words	3 times in 64186 words
Normalized frequency of the cut-off point	40 occurrences per million words	

These normalised raw frequency levels are set as a basis for selecting which clusters to analyse and, the next step then is to decide the length of the clusters to be analysed. Various arguments have been put forward for deciding the length of the clusters examined. Stubbs and Barth (2003: 269) describe 2-grams (e.g. *at the*) as predominantly preposition-determiner combinations, 3-grams (e.g. *at the end*) as function and content word combinations (often containing delexicalized content words), and 4-grams as having 'clear differences in topic and function' between different registers (e.g. *at the end of*). For Cortes (2004: 401), two-word clusters do not even qualify as what she terms 'lexical bundles', as she argues lexical bundles are frequent combinations of three or more words. She investigated four-word bundles because they "present a wider variety of structures and function to analyse" (Cortes, 2004:401).

Hyland (2008a, b) discusses the length of strings, and he states that four-word bundles appear to be the most often studied, as three-word bundles are extremely common and seem not to be interesting, whereas five- and six-grams are comparatively rare and often subsume shorter ones. Leedham (2011:159) also supports the pragmatic decision

on the length of clusters for analysis, based on the rationale that it can "provide sufficient, yet not excessive data for analysis" (also see Cortes, 2004; Hyland, 2008a, b).

The third criterion is that bundles have to occur in a specified number of files in the corpus, such as three to five texts (Biber and Barbieri, 2007), or 10% of texts (Hyland, 2008), in order to avoid the idiosyncrasies of individuals. Since the two corpora in this study consist of four texts written by each student at most, the occurrence has to be across at least five texts to ensure the language is not produced by "individual quirks" (Hyland, 2012).

The examples, as illustrated in Figures 5.8 and 5.9 below show two texts with the four-word clusters identified within them.

As it is known to us all, December 5th is International Volunteers Day. To be a volunteer is really a good and meaningful thing in that you can gain a lot from it, such as making new friends with others, putting what you have learnt into practice, to improve your skills and so forth.

When you are going to be a volunteer, you will do something helpful to others, so maybe you will make friends with those who you help and those who are volunteers just like you. When you go to the west of China to support that local education, you'll put what you have learnt in class into practice. I think it's a really exciting and interesting thing to use your knowledge to help rural area children. What's more, if you are a volunteer, you will improve your skills. For instance, thinking you are a Olympic Games volunteer, you are due to meet many people. Some of them will communicate with you, others may need your help. So you will improve your skills of communicating, ability to do things.

In a word, to be a volunteer is a really beneficial thing. If you join the volunteer team, you will surprisingly find you are not what you used to be, in other words, you will be a kinder, experienced, knowledgeable person.

Figure 5.8: The Year 1 written texts sample with extracted four-word clusters

Recently, there is a hot issue that is widely debated, the topic is: whether should we help strangers? In terms of this, I have something to say.

Indeed, these cases which the strangers cheat or hurt the one who have helped them before exist, and due to the widespread of media, these case are being magnified; thus causing a worldwide concern on themselves' safety but those phenomena are in some degree on the last part. In my view, I still think that we should help strangers, here is as follows.

For thousands of hundreds years, we have been taught acting nice to everyone. Thus, we win such a harmony (hamony) society, the development of society need a warm atmosphere too. If we concentrate too much on the bad aspect, we won't worry ourselves all the time, we thus cannot spare our time enjoying the beauty around us. It's not the true meaning of life at all, so giving our hands to these who need help. It is not only helping other, it helps ourselves too, the energy of warm can be transferred through our hands. Those repay their helpers badly can sooner or later shamed of what they did, apart from this, when we help strangers, we should more or less be aware (ware) of getting cheated and use common sense to tell them, after doing this, we definitely (definately) can gain ourselves satisfactory!

Figure 5.9: The Year 3 written texts sample with extracted four-word clusters

The quantitative analysis of four-word clusters was undertaken with Wordsmith Tools 6.0 (Scott, 2012) that identified and stored every four-word clusters in the two corpora. The programme analysed each text in the corpus, storing every cluster beginning with the first word of the text and advancing one word at a time. For example, the first sentence in figure 5.8 would have the following four-word sequences:

> *as it is known, it is known to, is known to us, known to us all, to us all December, us all December 5th, all December 5th is, December 5th is International, 5th is international Volunteers, is international Volunteers Day.*

In addition to this, orthographic word units are used and two contracted words are considered as one single word (*I'll try my best*); and lexical clusters that spanned a turn boundary or a punctuation mark were excluded (Biber et al., 1999). Thus the clusters are extracted by the Wordsmith tool 6.0 with the option 'stop at punctuation'. This is also one of the main reasons that I chose to use Wordsmith Tools over other corpus analytic

tools for this study, in addition to its user-friendliness. Overall, there are 1,790 four-word clusters in the Year 1 corpus and 918 four-word clusters in the Year 3 corpus.

Based on the extraction criteria, these clusters, at least, have to occur twice and appear in five texts. In addition, the clusters spanned over any punctuation were ruled out for analysis. Therefore, only 3 four- word clusters in the Year 1 text sample were chosen for analysis; these are: *is known to us, is international volunteer day, to be a volunteer*. Only 3 clusters in the Year 3 text sample were chosen for the analysis, which are: *should we help strangers, that we should help, we should help strangers*. They are underlined in the Figures 5.8 and 5.9.

In my study, the numbers of three-, four-, and five- word clusters in both corpora are listed in the table below:

Table 5.8: The numbers of three-, four-, and five- word clusters in both corpora

Sub-corpora	**Three-word clusters**	**Four -word clusters**	**Five -word clusters**
Year 1	499	158	65
Year 3	633	276	168

It is clear from the table above that four-word clusters are a manageable yet informative amount for analysis. The complete lists of four-word clusters extracted from both corpora can be found in Appendix 11.

As can be seen in Appendix 11, the four-word clusters are all continuous sequences and are used frequently in the corpora, regardless of their coherence as a single unit. The content-/topic-related bundles are included in the lists, too. Biber (2006) and Hyland (2012) conclude that the lexical bundles are made evident through corpus analysis software that retrieves multi-word units with specific frequency and distribution criteria and as a result, they are neither idiomatic nor, usually, complete grammatical units. Some examples from this study are *with the development of*, *is one of the*, *as far as I*, *it is said that*.

So far, after the first two phases, the total amount of identified formulaic language in this study, assessed by the by the number of formulaic strings and four-word clusters, is shown in table 5.9 below:

Table 5.9: The total numbers of formulaic strings and four-word clusters

Corpora	**Strings**	**clusters**
Year 1	210	158
Year 3	220	276
Total No.	430	434

It emerges that the total number of formulaic strings and four-word clusters in the Year 1 and Year 3 corpora are quite similar. These are 430 formulaic strings and 434 four-word clusters respectively (more about the differences in formulaic language in terms of year groups will be discussed in section 6.1.4). The Third Phase aims to implement the structural and functional classification of formulaic strings and four-word clusters identified in Phase 1 and 2.

Phase 3: The classification of identified formulaic language

The student identified formulaic strings and the four-word clusters were then categorised structurally, using Biber et al.'s (1999) classification system and grouped functionally, using an adapted version of Hyland's (2008a, b) categorisation system (see section 2.2.1). It is hoped that this will give more detailed insights into the formulaic language used in Chinese students' writing, and shed some light on the conceptualisation of formulaic language and related methodological issues from students' perspectives. Next, the analytic frameworks for structural and functional classification will be introduced in turn.

1) Structural classification of formulaic language

As Leedham (2011:118) points out, irrespective of whether a chunk is holistic, for example, a formulaic sequence, or simply frequent as an N-gram, it can be possible to syntactically classify them on the basis of the first part of the item, which is also the usual method adopted in the studies on structural classification of lexical clusters.

In order to compare the formulaic language identified by both approaches, both student-identified strings and four-word clusters which were classified according to their structural patterns. I followed some researchers (Baker and Chen, 2010; Biber, Conrad, and Cortes, 2004; Bal,2010; Leedham, 2011, Salazar,2011) and adapted Biber et al's (2004) classification system and grouped those structural patterns together as phrasal bundles, including NP-based, PP-based, VP-based bundles, adjective phrases (Adj. P) and adverbial phrases (Adv.P); and bundles that incorporate clauses (hereby clausal bundles) (see section 2.2.1.1). The classification followed here uses both Biber et al.'s academic writing patterns and the four broad categories shown in Figure 5.7. In the process, I first assigned the strings and clusters into the Biber et al.'s structural categories, i.e. sub-categories, and then grouped them together under each broad category.

Broad category	**Structural patterns**	**Examples**	
		Strings	**clusters**
clausal	(verb/adj.+) to-clause (fragment)	*Lucky enough to;* *to affect every one*	*Like to make friends;* *have volunteered to help;*
	as-clause (fragment)	*As it is said;* *as it is known;*	*As is known to;* *As our ancestors often;*
	if-clause (fragment)	*If you…, you will…*	*If you want to;*
	(NP+)(verb) that-clause	-[4]	*No doubt that we;* *that helping others can*
	1st/2nd person pronoun +dependent clause	*I hold the view that; I hope I'll…*	*I think it is;* *I hope I can;*
	Wh-clause	*Which strongly illustrated;* *What you should do is to;*	*Who have provided more;* *which is said to;* *When it comes to;*
Phrasal	NP-based	*The dragon boat festival;*	*One of the most;* *People who are in;*

[4] There are no examples found under the category (NP+) (verb) that-clause in the formulaic strings identified by the both year groups in this study. See section 6.1.3 for more detail discussion.

		the Qing Ming festival	
	VP-based	*Deal with;* *concentrate on*	*Is one of the;* *Is known to us*
	PP-based	*In addition; in the morning*	*At the same time;* *All over the world*
	Adv.p/Adj.p	*More than; now that;* *very much; most of*	*Very important for us;* *Important for us to*
Other structures	Ill-formed	*As well; and so forth*	*Not only can we;*
	Formulae (sayings, quotations, idioms)	*Where there is a river, there is a city;* *no pain, no gain;* *rose to lovers, smell in hands*	*As far as I;* *Far as I am;* *More and more (popular);*
	Templates	*For one thing…for another;* *Firstly…secondly…finally*	*Not only can we*
	Other Expressions	*The air is so fresh for us to breathe.* *Firstly…; secondly…; finally…*	*High and trees grow;*

Figure 5.10: Structural categories of formulaic language

This structural classification aims to provide a platform to compare the formulaic language identified by different ways and in other studies, rather than defining the necessary structural or grammatical features of formulaic language in a general sense. Also, it is not an exhaustive analytic framework, but it exclusively helps to describe the formulaic language identified in this study at both phrasal and clausal levels.

The first broad category is clausal and includes six sub-categories, which are (verb/adj.+) to-clause (fragment), as-clause (fragment), if-clause (fragment), (NP+) (verb) that-clause, 1st/2nd person pronoun +dependent clause, and Wh-clause. The 2004 study by Biber, Conrad and Cortes, used the bundles incorporating dependent clause fragments in addition to simple VP fragments. In my study, due to a large amount of data, especially of VPs, I put this type of formulaic language in an independent broad category at clausal

level. Strings and clusters identified in this broad category begin with a main clause (for example, *I want you to*) followed by a complementizer (such as *to, if*), or a WH-word introducing a dependent clause, for example, *which is said to*; other bundles in this broad category are dependent clauses or fragments beginning with a complementizer or subordinator, for instance, *I think it is, I hold the view that*.

At phrasal level, there are four sub-categories, i.e. VP-based, NP-based, PP-based, and Adj.P/ Adv.P. Previous research on the structural description of formulaic language (Biber, et al., 1999; Biber and Barbieri, 2007; Biber, Conrad, and Cortes, 2004; Chen and Baker, 2010; Bal, 2010; Leedham, 2011, Salazar, 2011) grouped the structural patterns into NP-based, PP-based and verb-phrase-based sub-categories and provided similar further classification under each sub-category. This provides a more detailed manual for the classification of identified strings and clusters. The detailed structural classification can be found in Appendices 12 and 13. The table below shows the further structural patterns in each sub-category.

Table 5.10: The further structural patterns in each sub-category

Sub-category	**Further structural patterns**	**Examples**	
		Strings	clusters
NP-based	NP+ of -phrase (fragment)	*All kinds of*	*The beginning of the*
	NP+ other post-modifier (fragment)	*A very place where hold*	*A time to plant*
	other NP	*Family quality time*	*The trees and grass*
PP-based	PP+ embedded of-phrase (fragment)	*For thousands of hundred years*	*With the development of*
	PP+ other post-modifier (fragment)	*Despite the fact that*	*Important traditional festival in*
	other PP	*With day and night*	*At the same time*
VP-based	copula be + NP/Adj.P	*Be full of*	*Is no doubt that*

	Anticipatory it+ VP/Adj.P+ complement clause	*It is necessary for*	-[5]
	passive VP (fragment)	*Born with courage*	*Has always been considered*
	pronoun/ NP+ BE	*Here are reasons*	*A volunteer is a*
	active VP (fragment)	*Feel sorry for*	*Want to be a*
Adj.P/ Adv.p		*Not familiar with*	*Important for us to*

In the table above, I add the last sub-category Adj.P/ Adv.P in the structural classification because usually it was put in the category of "others" in previous studies. In my study, I use the category of "other expressions" to refer to a different type of formulaic language.

Table 5.10 not only provides a detailed manual for the structural classification of identified formulaic language in the current study, but it is also a useful way to compare the smaller categories within or across studies when necessary. For example, Hyland and Tse (2005) adopted Biber et al.'s (1999) classification system and found that academic writing uses a high number of "NP+ post-modifier fragments", such as *the number of*, *the relationship between the*; "PP + embedded of fragments", such as *in terms of*, *on the basis of*, and "Anticipatory it fragments", for example, *it was found that* and *it should be noted that*.

In this study, as a new category in addition to the clausal and phrasal level description, the category of "other" mainly includes (a) ill-formed idioms; (b) formulae;(c) templates and (d) others.

Ill-formed idioms are those that are not syntactically well formed, thus they cannot be assigned to either of the previous two broad categories, and examples include like *as well*, *and so forth*, *last but not least*. Formulae refer to expressions such as sayings, proverbs, and similes, examples like *No pain, no gain*, *Nice to meet you!* Templates refer to the expressions that the writers used to frame a sentence or discourse. By using

[5] There are no examples of "Anticipatory it+ VP/Adj.P+ complement clause" in four-word clusters in this study.

expressions like *Firstly…Secondly…Thirdly*, for example, students can frame the paragraphs in a text. Also, they can use *for one thing…for another* to introduce their opinions in a sentence.

The last sub-category is "Other expressions", which includes some sentences and phrases or fragments that do not fit into any of the previously discussed categories. For instance, *hot issues vary from day to day*.

For the category of idiomatic phrases, I adopt the definition of idioms in a broader sense, according to Moon's 1998 study (see section 2.2.1.1). As a practical delimitation, I also used the Oxford learner's dictionary. Examples of this group include, *as well as, at the same time; as a result*.

As shown in Appendix 10, students also marked some individual words. There are 25 words identified by the students as formulaic language. All these 25 words were further divided into three sub-categories: 1) derivations or inflections of a word, such as *gratification, unfortunately, outstanding*. 2) Compound words, including closed forms and hyphenated forms, such as *outstanding*, *ex-wheeler*, and *short-term*. 3) Conjunction words, such as *meanwhile, and, but,* etc. The reason that the students considered the conjunction words to be formulaic language arguably lies in the fact that there are no explicit conjunction words in Chinese writing, while in English these are required as discourse devices to make the texts grammatical, cohesive and coherent. For the first two types of individual words, part of the word was already known by the students, and then they just added some changes based on the fixed rules. I am aware that this is a very unusual decision, as the main defining criterion for formulaic sequences is their combining of several words. Thus, these words were ruled out of this study.

2) Functional classification of formulaic language

O'Keeffe et al. (2007) propose the term "pragmatic integrity" to denote the pragmatically specialised roles that lexical chunks fulfil in discourse, a notion of functional adequacy that is independent of structural completeness. They argue that

> it is in pragmatic categories rather than syntactic or semantic ones that we are likely to find the reasons why many of the strings of words are so recurrent […] by pragmatic categories we mean the different ways of creating speaker meanings in context (O'Keeffe, et al., 2007:71).

According to Salazar (2011), it is necessary to link the lexical chunks to pragmatic categories to explain why certain structures are more frequent than others. This statement is in line with Biber and Barbieri's (2007) findings, that as one type of formulaic language, lexical bundles do serve important discourse functions related to the expression of stance, discourse organisation, and referential framing. However, the relationship between the form and function of formulaic language has generally been overlooked by the research from the psycholinguistic perspective (see section 2.2.1).

From a pedagogical perspective, researchers argue that the functional analysis of formulaic language is essential to its value as a teaching item (Biber and Barbieri, 2007; Hyland, 2008a, 2008b; Leedham, 2011; Salazar, 2011) (see sections 2.2 and 2.3).

To further describe the role of formulaic language in these written texts, a typology of discourse functions of formulaic language was proposed (see Figure 5. 8), which is based on previous research (Biber and Barbieri, 2007; Hyland, 2008). Following Halliday (1978) and Martin (2009), a natural relation is posited between the organisation of language and the organisation of social context, built up around the notion of three kinds of meaning, i.e. textual, interpersonal, and ideational. Martin (2009) specifies that interpersonal meaning is related to the enactment of social relations or social reality, referred to as "tenor" in his classification; ideational meaning is related to the construction of institutional activity, termed "field", and the textual meaning is related to the information flow across media, referred to as "mode". In the case of writing, the author needs to convey the subject matter or the topic (field) to the reader, through the written texts (mode); and the process of conveying realises the tenor, which indicates the interpersonal meaning.

Previous studies taking a functional approach usually analyse chunks using a taxonomy in three ways, along with the metafunctions identified in Halliday's SFG (Halliday, 1994;

Halliday and Matthiessen, 2004), although not using his terms “interpersonal”, “ideational” or “textual” metafunctions (Biber, Conrad, and Cortes, 2004; Biber, 2006; Hyland, 2008a, b; Pecorari, 2009). Biber, et al (1999) identified referential bundles, text organisers, stance bundles and interactional bundles. Cortes (2004) follows Biber et al.’s (1999) study and groups the last two functions together as interpersonal functions. For example, Hyland (2008) proposes a new typology of lexical bundles and collects them into broad three foci of research, text, and participants. These bundles can be fitted into the framework of field (research-oriented), tenor (participant-oriented), and mode (text-oriented). This study draws on both works, and as shown in Figure 8 below, the identified formulaic sequences were assigned to three main categories, textual, interpersonal and ideational functions.

The textual category, related to mode, is connected with the overall organisation of the text. Based on Hyland’s (2008a) study, the contrastive and resultative functions were further narrowed in my study and were substituted by additive, comparative and inferential respectively. Chunks in this category may mark inferential and causative relations between elements (regulative), organise the discourse (structuring), signal additive or contrastive links between elements (transition), or specify limiting conditions of the arguments (framing). Two new sub-categories were added: generalisation, which refers to the function that signals generally accepted facts or statements; and exemplification, which is usually used to give examples. The changes made under the textual function are due to the features of the student writing collected for this study, thus the topics of students’ writing are accessible from the real world. Students are unlikely to have any specialist knowledge of the area they were writing about, so their answers draw on anecdotes or personal experience. Moreover, this type of student writing is often described as “content-free writing”, as it does not display textual plurality, that is, there is no citation of other texts, and writing is a separate activity from reading (Horowitz, 1986; Leedham, 2011; Moore and Morton, 2005).

The interpersonal category, performed as tenor, focuses on the writer or reader of the text and expresses the attitude or assessment of the propositions. It is further divided into two aspects, ‘Engagement’ and ‘Stance’.

Engagement often involves an interactive relationship with the reader with directive addresses or advice; Stance mainly conveys the writer's attitude, responsibility, intention or evaluations. As for the category of ideation, referred to as field, it is concerned with real-world experiences and activities, which include location (time and place), description, action, and quantification (see section 6.1.1 for a more detailed explanation regarding the structural classification of formulaic strings and four-word clusters).

The last functional category is the ideational function, which helps the writer to structure their activities and experience of the real world (Hyland, 2008a: 18). It is the same as Halliday's ideational function. Formulaic language identified with ideational function is usually used to indicate place or time (location); or to describe measures, quantities, proportions, and changes (quantification), or to indicate quality, degree and existence (description); or to describe activities or experience (action).

Broad functional category	**sub-category**	**examples**	
		Strings	Clusters
1. Textual (mode)	Resultative	Now that	Tend to be
	Structuring	When it comes to	In my point of
	Generalization	As is known to all	It is said that
	Transition	On the one hand…on the other hand	It is not only
	Exemplification	For instance	-[6]
	Framing	In conclusion	To begin with
2. Interpersonal (tenor)	Stance	I firmly believe that	Is the most important
	Engagement	Maybe you know	If you want to
3. Ideational (field)	Location	Just around the corner	
	Description	Something helpful	The first time I
	Action	Pay respect to	I went to the
	Quantification	In some degree	A lot of things

[6] There are no examples of clusters served exemplification, location and conversational function in this study.

Conversational	Simple inquiry Conversational	What about my dreams?	-

Figure 5.11: The functional categories of formulaic language

Two issues regarding this functional classification need to be clarified. First, in the same way as the structural classification discussed in the previous section, this functional classification does not aim to provide an exhaustive typology for formulaic language, but rather to offer an exclusive descriptive account of the functions of identified formulaic language in this study. Second, there are some difficulties with a mono-functional classification (Leedham, 2011; Oakey, 2009; Salazar, 2011). Biber, Conrad and Cortes (2004) acknowledge that a single lexical bundle can serve multiple functions in different contexts. However, Hyland does not discuss how this is resolved in categorising the chunks in his studies. This lack of clarification makes Hyland's work difficult to replicate since the basis for categorization is not always clear. The usual solution proposed by researchers is to examine the concordances of potentially multifunctional bundles and classify them according to their most common use. For example, *at the same time*, can serve as the ideational function to describe the fact that real world events happen coincidently, or it can be an additive signal, to add new information or arguments by serving a textual function. However, in this study, the primary function of *at the same time* is additive. Keeping a primary single functional categorisation of each type is a pragmatic solution to a complex problem since classifying each token individually is not workable in a study on such a scale.

This section has delineated Phase 3 in the analytic procedure and discussed two taxonomies for categorising identified formulaic language in this study: one of these is based on their structural form and follows studies by Biber et al. (1999) and Biber, Conrad and Cortes (2004). The second taxonomy follows by Hyland's (2008a, b) studies and groups formulaic strings and four-word clusters according to their function; however, Hyland provides fewer examples and assumes that each chunk has a single dominant function. As a feasible decision made based on a pragmatic solution, I assign each type of formulaic strings and clusters to a single primary functional category.

Phase 4: Synthesis and comparison of formulaic strings and four-word clusters across groups

The final phase of the analytical process involved making researcher-led decisions on the inclusion and exclusion of sequences identified by these approaches as fully formulaic. I used the terms "core formulaic language" and "shared formulaic language" to describe the convergent use of formulaic strings and lexical clusters identified by two approaches. The term "core formulaic language" refers to the same or overlapping strings and clusters identified in one corpus; whereas the "share formulaic language refers to the same or overlapping strings and clusters in two corpora, including shared formulaic strings and four-word clusters respectively" (see section 6.1.5 and see Appendix 14). The analysis will then be led by an extended discussion of the core and shared formulaic language and their classification.

As one aim of the study is to use two different approaches to describing the formulaic language used by the Chinese university students in their English written texts, the final stage of the textual analysis synthesises and compares the formulaic language identified by two independent approaches, i.e. student-identified formulaic strings and four-word clusters extracted by corpus linguistics.

First, the analysis focuses on the discrepancies in the results from the two approaches, with regard to the typological, structural and functional distributions. It then takes a closer look at the shared formulaic language identified by both approaches in each sub-corpus. The list of shared formulaic language is presented and analysed in terms of structure and function (see section 6.1.3).

As it is known to us all, December 5th is International Volunteers Day. **To be a** volunteer **is** really a good and meaningful thing in that you can gain a lot from it, such as **making new friends with** others, putting **what you have learnt** into practice, to **improve your skills and so forth**.

When you are going to be a volunteer, you will do **something helpful** to others, so maybe you will make friends with **those who** you help and those who are volunteers just like you. When you go to the west of China to support that local education, you'll **put** what you have learnt in class **into practice**. **I think it's a** really exciting and interesting thing to use your knowledge to help rural area children. **What's more**, if you are a volunteer, **you will** improve your skills. **For**

instance, thinking you are a Olympic Games volunteer, **you are due to** meet many people. Some of them will communicate with you, others may need your help. So you will improve your skills of communicating, ability to do things.

In a word, **to be a volunteer** is a really beneficial thing. **If you** join the volunteer team, **you will** surprisingly find you are not what you used to be, **in other words**, you will be a kinder, experienced, knowledgeable person

Figure 5. 12: The student-identified formulaic strings (in bold) and four-word clusters (underlined) in the Year 1 written text sample

Figure 5.12 shows the 19 expressions that were identified by the students as formulaic strings in Bold and the 4 expressions extracted from the corpora as underlined Only two expressions were identified by the both approaches, i.e. *known to us all* and *to be a volunteer*.

Recently, **there is a hot issue** that is widely debated, the topic is: whether should we help strangers? **In terms of** this, I have something to say.

Indeed, these cases which the strangers cheat or hurt the one who have helped them before exist, and **due to** the widespread of media, these case are being magnified; thus causing a worldwide concern on themselves' safety but those phenomena are **in some degree** on the last part. **In my view**, I still think that we should help strangers, **here is as follows**.

For thousands of hundreds years, we have been taught acting nice to everyone. Thus, we win such a harmony (hamony) society, the development of society need a warm atmosphere too. If we concentrate too much on the bad aspect, we won't worry ourselves all the time, we thus cannot **spare our time enjoying** the beauty around us. It's not the true meaning of life at all, so giving our hands to these who need help. It is not only helping other, it helps ourselves too, the energy of warm can be transferred through our hands. Those repay their helpers badly can **sooner or later** shamed of what they did, apart from this, when we help strangers, we should **more or less** be aware (ware) of getting cheated and use common sense to tell them, after doing this, we definitely (definately) can gain ourselves satisfactory!

Figure 5.13: The student identified formulaic strings (in bold) and four-word clusters(underlined) in the Year 3 written text sample

Figure 5.13shows the 10 expressions identified by the student writers as formulaic strings and three instances of four-word clusters extracted from the corpora (*should we*

help strangers, *that we should help*, *we should help strangers)*. However, there are no expressions identified by both approaches (see section 6.1.3 for further discussion).

It is hoped that this phase of analysis will contribute to a further understanding of the relationship between frequency-based formulaic language (corpus-extracted four-word clusters) and formulaic language that is perceived as having processing advantages in students' psycholinguistic status (i.e. student identified formulaic strings) (see chapter 2 for more discussion on the speaker-internal and -external approaches to formulaic language).

Next, comparisons are made between the two sub-corpora, in other words, the examination of formulaic language use takes the group differences into account (see section 6.1.4). There is no definitive relationship between formulaic language use in English writing and language proficiency, but this analysis will shed some light on this issue in similar EFL settings.

After these comparisons, two new notions developed by the researcher through the analysis will be proposed in section 6.1.5: shared formulaic strings and clusters, and core formulaic language. Shared formulaic strings or clusters refer to those strings or clusters that are used by both Year 1 and Year 3 students, mostly in the same way Core formulaic language refers to the overlapping or similar formulaic strings and four-word clusters in one corpus, i.e. within one year group of students' written texts (see samples in Figures 5.12 and 5.13).

As the last step in Phase 4, the distribution of the structural and functional categories will be analysed and discussed. This will provide answers to research question 1c.

This section has delineated the analytic procedures for student written texts. The analysis includes four phases. The first two phases aim to refine the data for analysis: the operationalization of student-identified formulaic strings and corpus-extracted four-word clusters has been described in detail above. Phase 3 offers a structural and functional categorisation of formulaic language identified in this study. It is noted that this categorisation is exclusively developed for descriptive purpose in this study, rather than to provide a set of definitive features for the identification of formulaic language beyond the study. These three steps provide the premise for the final and important analytical phase: the synthesis and comparison of formulaic strings and four-word

clusters, with regard to types, structures, and functions. A detailed examination of shared formulaic (language identified by both approaches) in each sub-corpus will also be presented.

It is hoped that using these textual analytical procedures will provide insightful information to answer research question 1 and its sub-questions. At the same time, the findings from the textual analysis will finally be synthesised and compared with the results from students' interview data analysis, which will be discussed in the following section. The detailed textual analysis results will be presented in Chapter 6.

5.8.2 Interview analysis

The other type of data collected for this study is perceptual data obtained from the students' interviews (see Appendix 8 for interview examples). For the interview analysis, the first step involved transcribing the interviews and making decisions about what to include in the transcriptions, given the focus of the research on the content of the interviewees' responses. Thus, the transcription of the interviews only kept the plain words without any representation of the nonverbal features or prosody, as they were not considered a significant feature in the analysis.

Once the transcriptions had been completed and checked for consistency, they were uploaded in NVivo 10, which was used for the storage of the data and the coding. Nvivo 10 was chosen as it is designed as a platform for analysing all forms of unstructured data and enables interrogation of different sets of data using search, query, and visualisation tools. Both bottom-up and top-down approaches were used for the coding, which were related to the research focus on the repositioning of formulaic language, formulaic language learning, using and teaching. The top-down approach, or a priori coding, is used to analyse the data on the basis of pre-established code categories, whereas the bottom-up approach, i.e. a posteriori coding, refers to the code categories that emerge from the data by some processes of induction. There are a number of advantages to using a mixture of both approaches in the coding process, as some categories are pre-established while others are derived from the data. Pre-existing categories may not be enough to exhaust all the data, whereas the posterior approach is an attempt to develop

new categories and justify them all, which allows for flexibility in the research process (Wellington, et al., 2007).

In accordance with the research aims and questions, the data were initially coded based on the questions asked in the interviews with some pre-established code categories, based on the themes mentioned by the interviewees in the two rounds of interviews. For instance, for the first round of interviews, the initial codes for the transcriptions were: the importance of being nativelike, comparison between speaking and writing, recognition of formulaic language, teacher's emphasis on memorising formulaic language, formulaic language helps you to express yourself, where do you learn formulaic language, L1 influence on English writing, it is easy or difficult to have a good command of formulaic language in English, attitudes to the repeated use of formulaic language in written texts, attitudes to the repeated use of formulaic language in peers' written texts, and strategies used to deal with the repeated use of formulaic language in written texts. These codes are mainly established based on the interview questions from the schedules for two rounds of interviews (the top-down approach).

On the other hand, for the codes that emerged from the data (the bottom-up approach), and according to the frequency and depth with which these codes occurred in the interviews, the researcher revised and re-organised these codes as the research progressed. These code categories include 'formulaic language', 'native norms issues', 'examples of formulaic language', and 'contradictory points.' Further, the thematic codes for formulaic language were further developed into four aspects according to the data that emerged during the transcription phase, such as 'understanding the concept of formulaic language', 'formulaic language learning', 'using formulaic language' and 'formulaic language teaching'. In addition, the finalised codes are shown in Appendix 9.

As for the second round of interviews, the pre-established initial codes were 'six types of formulaic language', 'formulaic language teaching', and 'learning contexts'. Later on, these codes were divided into several sub-codes, which come from the transcripts. The code 'six types of formulaic language' for example, contains 'the understanding of six types of formulaic language', 'learning strategies used for each type of formulaic language' and 'learning preference'. All the finalised codes are in Appendix 9. Also, a

fellow researcher was asked to code two rounds of interviews with the code provided in order to examine and check any difficulties with the codes.

After the codes were finalised, the researcher completed individual reports on the in-depth data analysis under each code. The data from the interviews with two groups, Year 1 and Year 3, were analysed separately, and then compared at the end of each report. On the basis of these, the finalised code categories (see Appendix 9) integrated two rounds of interviews and are presented in the results section. The results section aims to achieve a rich interpretation of the data for the results of the interview analysis, as well as attempting to provide a holistic perspective on the themes which emerged from the interviews. At the same time, a comparative perspective will also be presented in order to examine the differences between the two groups. Also, it will suggest how the interview data relates to the research questions, which is a process that will contribute to the discussion in Chapter 7.

5.9 Research ethics

The informed consent of the participants was sought in accordance with the Ethics policy and guidelines of the University of Southampton. The full ethics permission and a positive risk assessment were obtained through the ERGO (Ethics and Research Governance Online) before the fieldwork started.

Both the consent form (see Appendix 1) and participant information sheets (See Appendix 2) were given to participants in the first meeting. The consent form was in both English and Chinese and the participant information sheet was in Chinese, in order to ensure that participants could fully understand the information. Participants also had the opportunity to ask questions of the researcher at this meeting and via email at later stages. At these meetings, the researcher emphasised again that the participants' involvement in the research was voluntary and that they could withdraw from the research at any stage. Moreover, it was stressed that participation was not related to their English course or other assessments in the university. The confidentiality of the data is protected through the use of pseudonyms. As the research was conducted at a

university site, this university's guidelines for staff and students were also adhered to (see section 4.3.2).

5.10 Triangulation of research instruments

Triangulation is used as one of the most efficient strategies to enhance the trustworthiness of the research (Merriam, 2009; Talmy and Richards, 2011; Suter, 2011). It encompasses the process of using multiple methods, data collection strategies and data sources to obtain a more comprehensive picture of what is being studied and to cross-check information (Gay, et al., 2006). As discussed in section 5.4, in one of the mixed-methods models, the triangulation design, the researcher simultaneously collects both quantitative and qualitative data, compares the results, and then uses those findings to see whether they validate each other (Fraenkel and Wallen, 2006). Through the use of triangulation, multiple data sources can be used to address each of the research questions.

To answer the first set of research questions, the exploratory research design was used, that is to say, formulaic language in the learners' written texts was first identified by the perceptions of the 12 informant writers and then the researcher used corpus methods to establish the use of formulaic language across all texts. By comparing the results from these speaker-internal and -external approaches, it is hoped that the study will provide an insightful description of formulaic language use by Chinese university students (see section 5.8.1).

As for the second set of research questions (RQ 2 and 2a), the triangulation design is employed to address the research problems. By examining the textual data collected through a quantitative method (compiling the student written corpora) and perceptual data collected through a qualitative method (semi-structured interviews), the student conceptualisation of formulaic language in English can be depicted. In addition, in order to improve the descriptive validity (Maxwell, 1992) and reliability (Dörnyei, 2007) of the perceptual data, the interviews were conducted in two rounds: the first round was conducted at the beginning of the fieldwork and the other one conducted nearly at the end of the fieldwork.

Regarding the last research questions, as it is exploratory and qualitative in nature, therefore it mainly relies on the perceptual data gathered from the interviews and relates to issues regarding the learning, use and teaching of formulaic language at the research site. The content is included from both rounds of interviews to ensure the reliability and validity of the data(see section 5.6.2).

5.11 Summary

This chapter has provided details of the research methodology for this study. After analysing the research questions, the researcher reviewed the main research methods extensively and provided a justification for the chosen research methodology, i. e, mixed methods. More precisely, the main research design is a convergent parallel design, drawing on both quantitative and qualitative research strands. The research design, data collection and data analysis of each strand will be kept relatively independent, with the exception of the exploratory design employed in analysing the first set of research questions. This chapter has provided detailed information about the selection of participants, researcher's role, fieldwork, research ethics and research instruments; followed by the explanations for data analysis procedures. The data analysis includes two parts: textual analysis and interview analysis. The textual analysis is further processed in four phases (see section 5.8.1). Finally, a discussion of the triangulation of research instruments used for this study is offered.

Chapter 6 Findings

This chapter will first present the initial findings of the data analysis based on the methodology described in Chapter 5. Based on this, the findings will be synthesised and discussed in relation to the three sets of the research questions in Chapter 7. In the following, analysis of students' written texts and interviews will be introduced in turn before arriving at a general discussion of the triangulation of the data and findings, as well as the limitations.

6.1 Formulaic language in student written corpora

This section will offer a description of the structure and function of formulaic language used in the students' texts written over the course of one semester, in order to shed some light on the first and second research questions. The first main research question regards formulaic language use in the students' English written texts, with four sub-questions including describing the main structures and functions of formulaic language, and the distributional relationship between the structural and functional categories, as well as identifying the differences in formulaic language use between the two groups of students (Year 1 and Year 3 respectively). The second set of research questions addresses the students' perception of formulaic language and examines the differences in the perceptions of the Year 1 and the Year 3 students (see section 1.2 or 5.1 for research questions). The answers to each research question and its sub-questions will be further discussed and presented in Chapter 7.

To answer the first research question, the textual data is analysed. With the collected written texts, I draw on two types of data. The first type of data is constituted by formulaic strings identified by the students in their own texts once they had finished writing; the second type of data is made up of the four-word clusters extracted from the two sub-corpora (i.e. the Year 1 and Year 3 corpus) based on the lexical bundle approach (Biber, et al., 1999; Biber, 2006; Hyland, 2008; Hyland, 2012). After analysing and comparing these two types of data, an answer to the question concerning the

description of formulaic language can be offered. As for the second research question, the perceptual data that students' identification of the formulaic strings offers can provide some explanation.

As clarified in section 5.8.1, two means of analysis were used in identifying formulaic language in these texts: One is based on psycholinguistic salience (the speaker-internal approach), that is, it uses student intuitions to identify and mark formulaic language in their texts (see figures 5.5 and 5.6); the other method relies on measuring the frequency of identical words clusters in the students' written texts (the speaker-external approach) by using a corpus linguistics approach. This section is going to present the analysis of the formulaic strings identified by the students.

In this section, these two types of identified formulaic language will be assigned to structural and functional categories and analysed. For the structure of the formulaic language, the linguistic structures will be examined, as these are assigned to three categories: phrasal, clausal and the 'other' structural category. In terms of the functional categories of formulaic language, on the basis of SFG theory, these are realised through the ideational (field), interpersonal (tenor) and textual categorisers (field). Finally, the findings will be synthesised and discussed. After the structural and functional classification of identified formulaic language, further comparisons will be the two identification approaches, as well as group differences, with regard to the quantity, structure, and function of formulaic language. This is followed by the examination of the convergences in the findings from these two approaches. This section finally will be ended up with a discussion on the distributional relationship between the structural and functional categories of the identified formulaic language.

This practice will ensure a rounded description of the formulaic language used by the learners in their written texts, as well as a basis for reflection on some issues in formulaic language in foreign language learning.

6.1.1 Structural classification: formulaic strings and four-word clusters

This section will first report the results and findings regarding the structural categorisation of formulaic strings identified by the students and four-word clusters extracted by the corpus linguistics approach, with a focus on an explanation of the structural and functional categorisation. At the same time, examples from the original texts will be presented to further illustrate the analysis and results.

As shown in Appendix 12, the categories of identified formulaic strings covered most of the structural correlates in section 5.8.1 at both clausal and phrasal level. Only one sub-category was not covered, which is "(NP+) (verb) that-clause" (see table 6.1), as no such example was found in the strings.

Tables 6.1 and 6.2 present the structural classification of formulaic strings and four-word clusters with the corresponding type frequency. Figure 6.1 shows the distribution of the types of three broad structural patterns of formulaic language in the Year 1 and Year 3 groups.

Table 6.1: The structural classification of formulaic strings with the type frequency

Structural patterns	**Types**	
	Year 1 (%)	**Year 3 (%)**
Clausal structures	**30 (14.29%)**	**13 (16.46%)**
(verb/adj.+) to-clause (fragment)	5	2
as-clause (fragment)	5	2
if-clause (fragment)	4	-
(NP+)(verb) that-clause	-	-
1st/2nd person pronoun +dependent clause	4	3
Wh-clause	12	6
Phrasal structures	**165 (78.57%)**	**187 (81.65%)**
NP-based	23	17
VP-based	96	123
PP-based	39	33
Adv.p/Adj.p	7	14
Other structures	**15 (7.14%)**	**20 (9.09%)**
Ill-formed	3	4
Formulae (sayings, quotations, idioms)	4	8
Templates	6	5

Others	2	3
Total	**210 (100%)**	**220 (100)**

Table 6.2: The structural classification of four-word clusters with the type frequency

Structural patterns	**Types**	
	Year 1 (%)	**Year 3 (%)**
Clausal structures	**26 (16.64%)**	**36 (13.05%)**
(verb/adj.+) to-clause (fragment)	12	19
as-clause (fragment)	1	2
if-clause (fragment)	1	-
(NP+) (verb) that-clause	-	4
1st/2nd person pronoun +dependent clause	2	1
Wh-clause	10	10
Phrasal structures	**128 (81.01%)**	**234 (84.57%)**
NP-based	39	92
VP-based	70	115
PP-based	18	25
Adv.p/Adj.p-based	1	2
Other structures	**4 (2.53%)**	**6 (2.18%)**
Ill-formed	-	-
Formulae (sayings, quotations, idioms)	-	-
Templates	1	3
Others	3	3
Total	**158 (100%)**	**276 (100%)**

It can be seen from the tables above, that overall, there are three broad types of both formulaic strings and four-word clusters identified in this study according to their structural characteristics, that is, clausal-based, phrasal-based and other structures. Next, the three broad structures with two types of identified formulaic language will be examined in turn.

6.1.1.1 Clausal-based

The clausal-based formulaic language in this study is comprised of six sub-categories. These are (verb/adj. +) to- clause, as-clause, if-clause. (NP+) (verb) that-clause, 1st/2nd person pronoun + dependent clause, and WH-clause.

1) (verb/adj. +) to- clause.

Formulaic language of this structure can be a simple to-clause or a to-clause preceded by a VP.

Most formulaic strings and four-word clusters with a simple to-clause are used to indicate the aim or purpose of the action, and most of them are with the present active voice in this study (1) (2)(3) (4).[7]

(1) *To be a* volunteer is a good way for us to learn from different people through different ways. (Year 3)
(2) *To memory this great person*, people celebrate on May5 in lunar (lunur), and it has lasts for about two thousand years. (Year 1)
(3) In my point of view, we'd better *try our best to* help strangers in a proper and reasonable way. (Year 3)
(4) But at the same time, we should also learn *to protect ourselves from* being harmed. (Year 3)

Some further items of formulaic language feature a predictive structure and control a to-clause to express ability (5) and likelihood.

(5) The people who will *be willing to read classics* become less and less. (Year 3)
(6) But nowadays, people *tend to be very cautious* of strangers, because they are afraid that helping others can sometimes bring them trouble. (Year 3)

2) as-clause

Almost all the formulaic language with as-clauses in this study is used by the writers to present a generalised proposition as an obvious and widely accepted fact or to introduce widely accepted knowledge or facts (7) (8). Sometimes, this type of formulaic language also expresses the writer's stance (9).

(7) *As is known to all* the Chinese, the Dragon Boat Festival, one of the most important traditional festivals. (Year 1)
(8) *As an old saying goes*, "give rose to others and the lasting fragrance will remain in you[r] hand". Therefore, we should help strangers. (Year 3)
(9) *We, as college student, should* let people to know the importance of reading classics. (Year 3)

3) 1st/2nd person pronoun +dependent clause

[7] All the examples extracted from the student written text by searching the identified formulaic strings or clusters in the Concordance programme in Wordsmith Tool 6.0.

Strings and clusters with this structure usually start with a first pronoun *I* and/or *we*, followed by a verb or VPs like *think*, *believe*, and *hold the view* that. Most of them are used to express the writer's opinion and perspectives (10) (11) (12). This type of structure has also been studied extensively with an exclusive focus on Chinese students (see Leedham, 2011)

(10) As a contemporary college student, *I hold the view that* it should be a day for us to memorise (memorize) him and study his sprit, but not only a day for us to entertain, though there are so many interesting activities. (Year 1)

(11) *I think it's* a really exciting and interesting thing to use your knowledge to help rural area children. (Year 1)

(12) At last, *we must be clear that* we should adopt appropriate ways to help others. (Year 3)

4) (NP+) (verb) that-clause

Not many formulaic strings or clusters were found with this structure in this study. There are only four types identified in the Year 3 clusters. These that-clauses introduced by an NP or simply by themselves usually serve to highlight a propositional statement (13) (14).

(13) There *is no doubt that* we should help strangers. (Year 3)

(14) While someone hold *that helping other is* helping ourselves, others suppose that giving help will bring us great trouble. (Year 3)

5) WH-clause and 6) If-clause

These two structural types seem do not appear in any studies on academic writing (see Bal, 2011; Hyland, 2008a, b; Leedham, 2011; Salazar, 2011,). However, they were identified as two sub-categories in the 2004 study by Biber, Conrad and Cortes on lexical bundles in university spoken and written registers. These types were also identified by two approaches used in the current study. These two types of formulaic language usually featured in the spoken register.

If-clauses in student written texts are mainly used to express engagement with the reader of their texts (15) (16) or the situation they imagined (17).

(15) *If you want to be loved*, love others. (year 1)

(16) *If you* can keep this hobby, *you will* have a good health. (Year 1)

(17) *If permitted* (permited), they can also drink a little wine----Xionghuang, which is said to protect people from illness and unluck. (Year 1)

For the WH-clause, it can be seen that these types of formulaic language take a certain proportion in strings and clusters across both year groups. In general, they are used to express the writer's exclamations and most cases are identified in Year 1 formulaic strings (18) (19) (20).

(18) *How exciting* game it is! (Year 1)

(19) *What a bad luck*! (Year 1)

(20) *What a lovely day*! (Year 1)

On the other hand, the WH-word is used to introduce a dependent clause. These are widely found in Year 1 clusters (21), and both formulaic strings and clusters in Year 3 student texts (22) (23).

(21) I'm willing to help people *who are in need* from the bottom of my hearts. (Year 1)

(22) Although there is an increasing number of others *who take advantage of* our sympathy we should try to give our trust to a stranger. (Year 3)

(23) As for TV news, it can help us know *what happened around*. (Year 1)

From tables 6.1 and 6.2, it can be noted that the identified formulaic strings or four-word clusters do not cover all the sub-categories of clausal-based formulaic language. This is partly because of the chosen methods for the identification. More explanation and discussion of the comparison of the two identification approaches will be found in section 6.1.3.

6.1.1.2 Phrasal-based

Phrasal-based formulaic language is divided into NP-based, prepositional phrase-based (PP-based), VP based (VP-based), and adjectival/adverbial phrase based (adj.P / adv.P -based) categories. In order to provide a detailed description of formulaic language use by the students, I further categorise phrasal based formulaic language into different sub-categories. Next, I will discuss them in turn.

1) NP- based

Tables 6.3 and 6.4 show the distribution of NP-based formulaic strings and four-word clusters in the Year 1 and Year 3 groups respectively.

Table 6.3: Yea1 Types of NP-based formulaic strings and four-word clusters in Yea1 texts

NP-based sub-structure	**Types**	
	Strings	**Clusters**
NP + of- phrase (fragment)	7	10
NP + other post modifiers	5	18
Other NP	11	11
Total	**23**	**39**

Table 6.4: Yea3 Types of NP-based formulaic strings and four-word clusters in Year 3 texts

NP-based sub-structure	**Types**	
	Strings	**Clusters**
NP + of- phrase (fragment)	4	28
NP + other post modifiers	2	19
Other NP	11	45
Total	**17**	**92**

Both formulaic strings and four-word clusters cover these three sub-categories of NP-based structure: NP+ of-phrase, NP+ other post-modifiers and other NP. However, compared with Tables 6.1 and 6.2, it is found that the NP-based is not the most common structure in both corpora. This result is not consistent with recent findings (Biber, et al, 1999; Byrd and Coxhead, 2010; Leedham, 2011; Salazar, 2011, Hyland, 2008a, b) and contradicts the view of academic writing as being "noun-centric" (Swales, 2008: v).

Most formulaic strings and clusters found in this study are used to describe events or people (*a money might world; best occupation; those who are in*) or actions (*the delight of giving; the passing from one generation to another*); to denote the quantity (*a kind of*; *a lot of things; tens of millions of*); or to mark location (*a traditional festival in; the hometown of*). Therefore, most NP-based formulaic language serves an ideational function (see section 6.1.2 for more detailed discussion).

2) PP- based

Similar to the NP-based formulaic language, the prepositional phrased-based formulaic strings and four-word clusters identified in this study were also further categorised into

three sub-structural patterns: PP+ of-phrase fragment; PP+ other post-modifiers; and other PP and fragments.

Tables 6.5 and 6.6 below show the types of PP-based formulaic strings and four-word clusters identified in this study in Year 1 and Year 3 group written texts respectively.

Table 6.5: Types of PP-based formulaic strings and four-word clusters in Year 1 texts

PP-based sub-structure	**Types**	
	Strings	**Clusters**
PP + of- phrase (fragment)	4	4
PP + other post modifiers	-	1
Other PP	35	13
Total	**39**	18

Table 6.6: Types of PP-based formulaic strings and four-word clusters in Year 3 texts

PP-based sub-structure	**Types**	
	Strings	**Clusters**
PP + of- phrase (fragment)	6	5
PP + other post modifiers	5	2
Other PP	22	18
Total	**33**	**25**

In the previous studies on chunks in academic writing (Biber, et al., 1999; Biber, Conrad, Cortes, 2004; Hyland, 2008a; Salazar, 2011; Leedham, 2011), it was found that most chunks have PP (fragment) structure, especially the structure PP+ of-phrase (fragment). In this study, on the whole, formulaic language with PP structure does not constitute the greatest proportion. Within the sub-structure in PP-based formulaic language, most identified formulaic strings and clusters have PP structure, followed by PP+ of-phrase (fragment).

In the current study, strings and clusters with other PP structures are mostly used to refer to different sections of the text (24) (25); to indicate a location (26), describe a situation (27), to exemplify (28) (29) or to indicate a time (30).

(24) *Above all*, the Qing Ming festival is a time to memorise our ancestors. (Year 3)

(25) *In conclusion*, there are others traditional festival like the Dragon Boat Festival we should make every effort to protect this kind of traditional

culture. (Year 1)

(26) *In the ancient China*, there was a preeminent (perminent) poet, Quyuan. (Year 1)

(27) Even though, they were mummified *with day and night*, always they possess a big smell (smile) to the victims. (Year 3)

(28) Taking experience *for example*, when I came to this school, I joined the English Volunteer Union, and did a lot of interesting things, among which one thing left a deep impression on me. (Year 1)

(29) As for our university students, we should take the torch of reading classics, and pass the tradition to the next generation, *for instance*, we can take some activities to tell people the importance of reading classics. (Year 3)

(30) Because we feel great joy when we are trusted and needed by others. But *at the same time*, we should also learn to protect ourselves from being harmed. (Year 3)

As for the strings and clusters with PP+ of-phrase fragment structure, mostly they are used to add more information to the texts (31) (32) (33) (34) (35) (36).

(31) It can help me improve my ability and, *on the other hand*, it can take me into society earlier. (Year 1)

(32) *In my opinion*, Chinese traditional festivals are our roots and souls, which can't be replaced by western cultures. (Year 1)

(33) *In contrast to* the sadness of the tomb sweepers, people also enjoy the hope of spring on this day. (Year 3)

(34) It makes we learn to be persistent *in spite of* difficulties, and it tells me what's the true meaning of our life, the most important, it leads me to the way to happiness and fulfilling. (Year 1)

(35) Beijing saw dramatic socio-economic development, especially *in terms of* resident's income, medical care and housing. (Year 3)

(36) They write some novels for fun *instead of* for cultural value and so on (year 3).

For the strings and clusters with the structure PP+ other post-modifier, they are usually extended from a PP (fragment) to another sentence or phrase (37) (38) to indicate a location or describe an event or action.

(37) *In China it is* also a festival to honour Qu Yuan, a citizen of country Chu, which was considered to be a poet and a national hero for his love to his

country. (Year 1)

(38) At last, I have to say every year there are so many new books *for us to read*, it is difficult for us to pick up the suitable book. (Year 3)

3) VP-based

From Tables 6.1 and 6.2, it can be seen that formulaic strings and four-word clusters with VP-based structures are mostly identified in this study. In order to gain insight into the use of formulaic language by the students in this study, I then put those VP-based strings and clusters into the following sub-structure: 1) copula be + NP/Adj.P; 2) Anticipatory it+ VP/Adj.P+ complement clause; 3) passive VP (fragment); 4) pronoun/ NP+ BE; 5) active VP (fragment).

Tables 6.7 and 6.8 below show the types of the VP-based formulaic strings and clusters in Year 1 and Year 3 written corpora respectively.

Table 6.7: Types of the VP-based formulaic strings and clusters in Year 1 written corpus

VP-based sub-structure	**Types**	
	Strings	**Clusters**
copula be + NP/Adj. P	2	12
Anticipatory it+ VP/Adj. P+ complement clause	3	-
passive VP (fragment)	6	-
pronoun/ NP+ BE	15	27
active VP (fragment)	70	31
Total	**96**	**70**

Table 6.8: Types of the VP-based formulaic strings and clusters in Year 3 written corpus

VP-based sub-structure	**Types**	
	Strings	**Clusters**
copula be + NP/Adj. P	9	16
Anticipatory it+ VP/Adj. P+ complement clause	7	-
passive VP (fragment)	8	1
pronoun/ NP+ BE	17	35
active VP (fragment)	82	63
Total	**123**	**115**

Copula be+ NP/adj.P, formulaic strings and cluster with this structure are mostly used to express writer's stance (39) (40) (41); or describe the event or action (42) (43).

(39) Classic books are *of vital importance to* one's growth. (Year 3)

(40) Reading classics *is very important to* people, we should notice this point and

repeal more people to read classics. (Year 3)

(41) *It is a good way* to deliver our culture by celebrating our traditional festival. (Year 1)

(42) My college *is far away from* my home so I often miss my parents. (Year 1).

(43) Sometimes, it seems so hard for us to understand what truly the author want to tell us. Thus, more and more people *are tired of* reading classics. (Year 3)

There are no clusters found in this study with structure anticipatory it+ VP/Adj. P+ complement clause, partly because one of the cluster extraction operationalization criteria was to choose stop at punctuation in the Wordsmith Tool 6.0. This structure is typically found after punctuation and is thus discarded in the cluster extraction programme (44).

(44) On the contrary, surf-online occupies much of the spare time, *it's no doubt that* there are many factors accounting for the phenomenon. (Year 3)

The majority of formulaic strings with this structure are employed by the writers in the appraisal of importance (45) (46), affirmation (47), or usefulness (48).

(45) Recently, the issue on whether *it is important to* read classics has aroused hot attention in public. (Year 3)

(46) I think reading classics *is of great importance to* a person's growing. (Year 3)

(47) So *it is necessary for* us to know that we should try our best to help others who are in need. (Year 1)

(48) *It is no doubt that* the Dragon Boat Festival is an important and meaningful festival. (Year 1)

Passive VP (fragment), as one of the techniques that academic writers usually employ to depersonalize, the passive VP (or fragment) structure was not significantly used in students' written texts. There are two tentative explanations for this: One is that the students' written texts collected in this study are nowhere near the level of academic scientific writing required; the other one lies the nature of the writing task itself, as most of the writing tasks that students were required to complete in this study were content-free (see Chapter 5 for more discussion) and based on personal accounts. Therefore, they did not need to cite others' work or data, on the contrary, they wrote about their personal stories or experience in the texts. Some of the strings and clusters using a

passive structure are still associated with real-world events (49) (50), rather than indicating of locative or logical relations (Hyland, 2008a; Salazar, 2011).

(49) They will realize it's wrong to take it for granted that they *are born to* live this kind of lives and they will value it. (Year 1)

(50) In the high pace of modern society, we must *be equipped with* knowledge, thoughts and inner peace. (Year 3)

Formulaic strings and clusters with the structure Pronoun/ NP+ be, are mostly "there be" sentences (51) (52) (53), or start with a personal pronoun to express the writer's stance (54).

(51) *There are* bad phenomena that make use of the warm-heated persons. (Year 3)

(52) In different areas, *there might be* other activities to observe the festival. (Year 1)

(53) *There is no doubt that* there are many people in the world need our help. (Year 1)

(54) *It is a fine time* to go out and to appreciate the beautiful scenes of nature during the festival. (Year 3)

Most VP-based formulaic strings and four-word clusters have an active VP structure. They are mostly used to describe actions. A substantial amount of these VPs contain phrasal verbs (*keep…from; fall on; concentrate on; call on; join in*).

3) Adjectival and adverbial phrase based

Identified formulaic strings and clusters falling in this category are adjectival phrases and adverbial phrases or fragments. They are used to compare (55) (56); or index the text (57) (58); or to give a personal appraisal (59).

(55) *More important*, it is a period to honour and to pay respect to one's deceased ancestors of family members. (Year 3)

(56) With the social's developing, there are *more and more* people participating in this activity. (Year 1)

(57) *Last but not least*, people should really remember and learn the spirit of Qu Yuan. (Year 1)

(58) *First and foremost*, the development of information technology makes it possible for people to watch various TV series which are composed on the

basis of famous novel. (Year3)

(59) In fact, it's very *important for us to* do more exercises. (Year 1)

6.1.1.3 Other structures

The third broad category includes all other formulaic strings and four-word clusters that do not fit into the previously described categories. Furthermore, these strings and clusters are divided into ill-formed, formulae, templates and other expressions, as shown in Tables 6.9 and 6.10.

Table 6.9: Types of other formulaic strings and clusters in Year 1 written corpus

other sub-structure	**Types**	
	Strings	**Clusters**
Ill-formed	3	-
Formulae	4	-
template	6	1
other expressions	2	3
Total	**15**	**4**

Table 6.10: Types of other formulaic strings and clusters in Year 3 written corpus

other sub-structure	**Types**	
	Strings	**Clusters**
Ill-formed	4	-
Formulae	8	-
template	5	3
Other expressions	3	3
Total	**20**	**6**

Ill-formed

The ill-formed formulaic strings and clusters refer to those that are not syntactically well-formed (Moon, 1998). As Moon (1998: 80-81) noted that these expressions 'cannot be parsed according to normal syntactic rules and they are non-compositional'; and usually these ill-formed expressions 'arise from odd phrase structures, ellipsis or inflections or from an archaic mood'. Examples include: *and so forth; what's more; as well; head over heels; as well as; few and far between*.

Formulae

Formulae include sayings, quotations, and proverbs. There are several types of intelligible formulae used by students:

1) original formulae: for example, the expression *no pain, no gain*; which is an exercise motto and is often used to expression meaning that greater value rewards for the price of hard and painful work[8];

2) structure-kept expressions: for example, *where there is [a]river, there is a city*, which actually comes from the proverb *where there is a will, there is a way,* in which the structure remains and the content is changed;

3) content-kept only expressions: where only the literal meaning is kept, for instance, formulae used by students "*The rose is in her hands, the flavour in mine"*, or "*rose to lover, smell in hands*" is originally from the sentence "*the fragrance always stays in the hand that gives the rose*, which comes from a poet[9].

However, there are other formulae which can be very difficult for readers from outside the students' socio-cultural backgrounds. For example, "where there is a book, there is gold". The structure of this sentence comes from the proverb "where there is a will, there is a way", whereas the content is translated from ancient Chinese literature "书中自有黄金屋 (*shu zhong zi you huang jin wu*)", which means reading can change his/her life into a wealthy one.

Apart from these, students tend to use some rhetorical devices, including metaphors (60), similes (61) and analogies (62). However, mostly, these are used inappropriately.

(60) *Like two stars, crashed together*. (Year 3)

(61) When we feel sorry for justice, we had to admit that *greed is the eternal soil of violence*. (Year 3)

(62) *The food is to Chinese people what the turkey to the Americans*. (Year 1)

For example, the sentence 'like two stars, crashed together' is not grammatical in terms of syntactic structure. In terms of meaning, the writer aims to describe two people

[8] It came into prominence after 1982 when actress Jane Fonda began to produce a series of aerobics workout videos. In these videos, Fonda would use "No pain, no gain" and "Feel the burn" as catchphrases for the concept of working out past the point of experiencing muscle aches.

[9] The fragrance always stays in the hand that gives the rose. Hada Bejar, 1640-1689, British Playwright poet

falling in love with each other deeply in a short time. In Chinese, this meaning can be realised by formulaic expression *tian lei gou dong di huo* (天雷勾动地火), which is a metaphor to describe the love between two people is so vigorous and powerful that just like the energy when two planet clash. In English, the informal expression '*get on like a house on fire*' is often used to describe such situation, It can be surmised that the student writers recognised Example (60) as formulaic language because they knew their formulaic equivalent in Chinese but not enough knowledge in English.

Example (61)is grammatical, as this expression was modified by the student from a Biblical quotation *'Greed is the root of all evil*'. However, it does not make sense in the sentence, as we cannot see any logical relation between the two parts in the sentence. It is a common rhetoric technique in Chinese writing that writers use quotations and motto in the text to show their knowledge of literature. In this case, the student uses this formulaic expression in his/her text but without clear meaning or logical connection in the sentence.

In Example (62), the writer uses an analogic structure to illustrate the relationship between a food and a country. Students probably identified this expression as formulaic language because this structure of analogy is very popular in the English textbook used*: A is to B what C is to D*. it is often used to transfer or convey meaning from the first part (*A is to B*) of the sentence to the second part of the sentence (*C is to D*). In terms of structure, usually 'as' instead of 'what' is used in this kind of analogic sentence, so it should be '*Th[is] food is to Chinese people <u>as</u> the turkey to the Americans*'. In terms of meaning, the food in the expression should be replaced with a specific type of food, for example, pork, to form an equal and similar relation as shown in the second part of the sentence (*the turkey to the Americans*). In this sense, the student has the awareness to apply an analogy in their text use the formulaic structure they have learned from the texts, however, the grammaticality and appropriateness still remained in question.

<u>Templates</u>

Formulaic strings and four-word clusters in this category take two sub-structures: macro templates are used to list the points across paragraph(s) in their exposition texts (63)

(64); micro templates are usually employed by writers to help them to structure the sentence (65) (66).

(63) In my opinion, there are some advantages for volunteering to help others. The details are as follows.
Firstly, everyone has the tendency to help others…
Secondly, volunteering can provide us an opportunity to exercise ourselves…
Last but not least, volunteering can also be a stage to realize your one value...
In a conclusion, volunteering benefits both you and the other. So just give your hand to other when someone is in need. (Year 1)

(64) *To begin with*, the economy increases fast, most of modern people are under great pressure about how to make more money, they try their best to make more money...*Then*, nowadays, the soap TV series are like the mushrooms after rain about classics... *At last*, there often have simple revised copies about classics, they are often easier to understand, and people would like to read them. (Year 3)

(65) *Only when* we hold the concept that do good deed not to get reward *can we* really feel this simple happiness from helping people. (Year 3)

(66) The Dragon Boat Festival is *not only* a festival to eat food or drive boat, *but also* a chance that we can think about the history. (Year 1)

Due to the limitations of the corpus linguistic approach, such as the length and word spans in clusters, the macro templates were only those recognised by the students and considered as formulaic strings (see section 5.10 and section 6.1.3 for discussion).

Other expressions

The rest of the formulaic strings and clusters fall into this final category, because either they do not fit in any of these categories discussed above, or they are meaningless for the analysis due to the limitations of the methods employed. Example (67) below, shows that this sentence was identified as a formulaic string by a student writer, but it can be considered as a case of "individual idiosyncrasies" (Wray, 2002); or as illustrated in example (68) and (69), the cluster *"his country very much*" strides over two multiword units with complete syntactic structures and semantic meaning, i.e. *he liked his country* and *very much and he*. This is a very unusual case in larger corpus studies. The small size

of the corpora in this study seems to result in a tendency for clusters to be more fragmental.

(67) *Hot issues vary from day to day*, however, some of them are not fresh at all, they just pop up with the development of society. (Year 3)

(68) He liked *his country very much* and he wanted do what he could to make all people feel happy and country strong enough to resist being invaded. (Year1)

(69) The Qing Ming festival is also a time to plant trees, for the survival rate of sapling is *high and trees grow* fast later. (Year3)

In this section, the three broad structural categories, together with each sub-structural category of formulaic strings and four-word clusters, are discussed and illustrated with examples. It builds the foundation for the later analysis and comparison of two identification approaches and differences across year groups.

6.1.2 Functional classification: formulaic strings and four-word clusters

As discussed in section 5.8.1, the functional classification of formulaic strings and four-word clusters used in this study follows studies conducted by Biber, Conrad and Cortes (2004), Halliday (1994), and Hyland (2008a, b) on the functional categorisation of formulaic chunks or lexical bundles. In order to achieve a feasible classification procedure, the researcher admits that some formulaic strings and four-word clusters identified in this study serve more than one function (see section 5.8.1 for more discussion). However, the functional classification used here is based on the primary or the most common function that a formulaic string and four-word cluster served. This section is going to present the results and findings of the functional categorisation of the strings and clusters, and some examples will be extracted through the Concordance programme in Wordsmith where necessary and presented as illustrations of this.

Tables 6.11 and 6.12 present the functional classification of formulaic strings and four-word clusters with the corresponding type frequency in Year 1 and Year 3 corpora.

Table 6.11: The functional classification of formulaic strings with the corresponding type frequency

sub-category	**Types**

	Year 1 (%)	Year 3 (%)
Textual function	**55 (26.19%)**	**55 (25%)**
Resultative	12	9
Structuring	5	10
Generalization	5	3
Transition	21	18
Exemplification	4	2
Framing	8	13
Interpersonal function	**39 (18.57%)**	**29 (13.18%)**
Stance	31	28
Engagement	8	1
Ideational function	**114 (54.29%)**	**136 (61.82%)**
Location	12	7
Description	34	49
Action	61	71
Quantification	7	9
Conversational function	**2 (0.95%)**	-
Total	**210 (100%)**	**220 (%)**

Table 6.12: The functional classification of four-word clusters with the corresponding type frequency

sub-category	**Types**	
	Year 1 (%)	Year 3 (%)
Textual function	**16 (10.13%)**	**25 (9.06%)**
Resultative	2	3
Structuring	-	-
Generalization	2	2
Transition	8	14
Exemplification	-	-
Framing	4	6
Interpersonal function	**14 (8.86%)**	**25 (9.06%)**
Stance	12	24
Engagement	2	1
Ideational function	**128 (81.01%)**	**226 (81.88%)**
Location	9	31
Description	78	127
Action	31	57
Quantification	10	11

Conversational function	-	-
Total	**158 (100%)**	**276 (100%)**

From both the tables above, it can be seen that the formulaic strings and four-word clusters identified in this study mainly serve three broad functions: textual, interpersonal and ideational. There are two exceptional cases that were identified by students as formulaic strings which undertake the conversational function in a written corpus. Next, these functional categories will be discussed in detail with examples.

6.1.2.1 Textual function

There are six further functional categories under the textual function, these are resultative, structuring, generalisation, transition, exemplification and framing, which are identified in the student written texts. The functions of generalisation and exemplification are added to Hyland's (2008a) original four sub-categories in the text-oriented lexical bundles in the academic writing. Next, these sub-categories of the textual function will be discussed in detail and illustrated with examples.

Tables 6.13 and 6.14 show the types of formulaic strings and four-word clusters with textual function in the Year 1 and Year 3 corpora respectively.

Table 6.13: Types of formulaic strings and four-word clusters with textual function in Year 1 corpus

Textual sub-function	**Types**	
	Strings	**Clusters**
Resultative	12	2
Structuring	5	-
Generalisation	5	2
Transition	21	8
Exemplification	4	**-**
Framing	8	4
Total	**55**	**16**

Table 6.14: Types of formulaic strings and four-word clusters with textual function in Year 3 corpus

Textual sub-function	**Types**	
	Strings	**Clusters**
Resultative	9	3
Structuring	10	-

Generalisation	3	2
Transition	18	14
Exemplification	2	-
Framing	13	6
Total	**55**	**25**

Resultative

The strings and clusters serving a resultative function in the texts are mainly used to indicate two types of links: inferential and causative relation. The former one is to signal inferences and conclusions drawn from the preceding argument or reasoning (70) (71); the causative markers are used to highlight cause-and-effect relationships (72) (73).

(70) *Now that* there are so many people who registered volunteers, what are the benefits of volunteering on earth? (Year 1)

(71) However, nothing can be changed, I had *no choice but to* learn English. (Year 1)

(72) *As a result*, people gradually form the habit of not reading classics, especially without teacher's and parents' demand. (Year 3)

(73) We can learn more from the classics, *so that we can* make use of some important knowledge. (Year 3)

Structuring

The structuring function is performed through text-reflexive markers that help organize stretches of discourse (74) (75).

(74) *In a word*, the dragon boat festival is an important and significant tradition in our China. (Year 1)

(75) *In conclusion*, helping strangers is one of the traditional Chinese virtues, and we ought to pass it on to our generation. (Year 3)

(76) The film Pretty Woman tells us a modern story about how a Cinderella turns into a princess…

Firstly, a smart mind makes a woman more beautiful...
Next, a sense of humour makes for a woman's success...
Finally, true love cannot be measured by money…
Even though not every girl can become a princess, they have chances and rights to seek true love. When (true love) it actually comes, they will become princesses of their own. (Year 3)

In this study, the majority of structuring formulaic strings and cluster are with prepositional phrases (74) (75) or in the form of a template (76).

Generalisation

As the writing tasks in this study require exposition writing and are mainly based on real world knowledge, common sense or personal experience, one of the two new added categories is the generalisation function. This function is to signal generally accepted facts or statements, and is usually realised by as-clausal structures (77) or anticipatory it structures (78).

(77) *As is known to us all*, the Dragon Boat Festival is one of the most famous and important traditional festivals in China. (Year 1)

(78) *It is widely acknowledged that* reading acts as a significant role during our growth, especially the classics composed by famous authors. (Year 3)

Transition

The function of the transition is to establish the additive (79) (80) or contrastive (81) (82) links in the texts. The transition function can be realised by different structures.

(79) *I firmly believe that* there is still true love between people, although there are also many deceptions and contradictions at the same time. (Year 1)

(80) It makes we learn to be persistent *in spite of* difficulties, and it tells me what's the true meaning of our life. (Year 1)

(81) In the north, because of lack of water, it's hard to hold the dragon boat race and the celebration is simple *comparing with* it in the south. (Year 1)

(82) *More important*, it is a period to honour and pay respect to ancestors and family members. (Year 3)

Exemplification

This the other new function identified in the formulaic language in this study, which is used to further explain the writers' viewpoints (83) (84) (85). There are no clusters with exemplification identified by the corpus linguistics approach in this study.

(83) To be a volunteer is really a good and meaningful thing in that you can gain a lot from it, such as making new friends with others, putting what you have learnt into practice, to improve your skills *and so forth*. (Year 1)

(84) What's more, if you are a volunteer, you will improve your skills. *For instance*, thinking you are an Olympic Games volunteer, you are due to meet many people. (Year 1)

(85) Moreover, we can organise some interesting activities, concerning the

reading classics. *For example*, we can persuade student to join the reading classics competition, and award the final winner. (Year 3)

Framing

The framing function is used to situate arguments by specifying the limiting conditions, and most of them are with prepositional (86) (87) and VPs (88) (89).

(86) *As far as I'm concerned*, volunteering to help those in need can give us three benefits. (Year 1)

(87) *In my point of view*, reading classics is indeed of great importance to us everyone. (Year 3)

(88) *It is said that*, by this way, people can get rid of disasters and diseases and bring about fortune. (Year 3)

(89) *As for* interest, sports must be my favourite. (Year 1)

The writers used this textual function formulaic language to try to lend coherence to their writing and to connect, clarify, contextualise, and exemplify their ideas. Although, sometimes, they made mistakes or used some formulaic language inappropriately (77) (87), at the same time, it is clear that the nature of the writing tasks also decides the main functions of the formulaic language identified in this study, which are different from those in academic writing (Hyland, 2008a, b; Leedham, 2011; Salazar, 2011). For example, there are two new sub-functions identified in this study: exemplification and generalisation, which are not normally discussed in academic formulaic sequences or lexical bundles.

6.1.2.2 Interpersonal function

Tables 6.15 and 6.16 display the types of formulaic strings and four-word clusters with interpersonal functions in Year 1 and Year 3 corpora.

Table 6.15: Types of formulaic strings and four-word clusters with interpersonal function in Year 1 corpus

interpersonal sub-function	**Types**	
	Strings	**Clusters**
Stance	31	12
Engagement	8	2
Total	**39**	**14**

Table 6.16: Types of formulaic strings and four-word clusters with interpersonal function in Year 3 corpus

Sub-interpersonal function	**Types**	
	Strings	**Clusters**
Stance	28	24
Engagement	1	1
Total	**29**	**25**

This functional category corresponds to the dialogic interaction between the participants in the text: the writer and the reader. This formulaic language is termed by Hyland (2005) as "participant-oriented bundles". By expressing epistemic, evaluative and directive meanings, these interpersonal expressions help writer convey their attitudes towards their assertion and establish the appropriate relationship with their readers. They are further divided into two sub-functions: formulaic language with stance features and those with engagement features.

Stance

In accordance with previous studies, it is also found in this study that stance marker are linguistic devices that carry meanings such as importance (90) and certainty (91), (see also Cortes, 2004; Salazar, 2011). In addition to this, the writers in this study seem to especially use the first plural personal pronoun we and modal verbs, for instance, *we should* (92) (93), to express the responsibility that they feel they bear.

(90) It *is the most important* sacrifice holiday, it is the day of worshiping their ancestors and the grave. (Year 3)

(91) There *is no doubt that* everyone has ability to love and to be lovable. (Year 3)

(92) *We should try our* best to care those who are needed. (Year 1)

(93) *We should pay more* attention to our own festival and protect our cultures well. (Year 1)

Engagement

The strings and clusters are used to engage the reader in the text and usually address readers directly (94) (95).

(94) For example, when you find someone being robbed, *what you should do* is

to call the police rather than chase the robber. (Year 3)

(95) *I think we should* keep the habit and teach it to our children. (Year 3)

Furthermore, a series of studies found that in academic writing, writers usually employed two techniques to realise depersonalisation: one is to use impersonal structures, for instance, anticipatory construction; the other is to adopt an indirect approach, which is known as "hedging" (Hyland, 2008a, b; O'Keeffe, et al., 2007; Salazar, 2011;). In this way, they can achieve the objectivity and precision on which academic writing places considerable emphasis.

However, in this current study, it is clear that students are more expressive with their attitudes and evaluations. As can be seen in both the tables above, there are far more stance markers than engagement markers in both year groups identified by both approaches.

For both stance and engagement markers, rather using an indirect approach or depersonalisation techniques, students in this study seemed to adopt a direct and personal approach, with a special preference for using formulae. For example, the structure first and/or second person pronoun+ dependent clause (fragment) is frequently used to express the writer's stance (96) (97); the writers also use exclamation, with the WH-clause to express their feelings and evaluation directly (98) (99); In addition, formulae are often used to express their evaluations and attitudes towards some events, to realise the persuasion (100) (101). These are very unusual in academic writing.

(96) *If you want to* improve yourself, you can join it. (Year 1)

(97) *We must be clear that* we should adopt appropriate ways to help others. (Year 3)

(98) *What a bad luck*! (Year 1)

(99) *What a busy day*! (Year 1)

(101) *Do onto others as we would have them done onto us*. (Year 3)

(102) *Help other is help ourselves*. (Year 3)

6.1.2.3 Ideational function

Based on Hyland's (2008a) study, some modifications were made to the sub-categories of ideational function in this study. This broad functional category comprises four subcategories: location, description, action and quantification. Tables 6.17 and 6.18 below show the types of strings and clusters that serve ideational function across year groups.

Table 6.17: Types of formulaic strings and four-word clusters with ideational function in Year 1 corpus

Ideational sub-function	**Types**	
	Strings	**Clusters**
Location	12	9
Description	34	78
Action	61	31
Quantification	7	10
Total	**113**	**128**

Table 6.18: Types of formulaic strings and four-word clusters with ideational function in Year 3 corpus

Ideational sub-function	**Types**	
	Strings	**Clusters**
Location	7	31
Description	49	127
Action	71	57
Quantification	9	11
Total	**136**	**226**

The formulaic strings and four-word clusters identified in this study with the ideational function mainly serve the following four sub-function in the written texts: location, description, action, and quantification, this is on the basis of the taxonomy used in the previous studies on the academic lexical bundles (Hyland, 2008a, b; Leedham, 2011; Salazar, 2011),

<u>Location</u>

Strings and clusters identified in the category of "location", usually serve to indicate time or place. Most of them have a PP structure (102) (103); some of them take the form of NP (104).

(102) *At the very point*, one of the official Jie Zitui cut a piece of flesh off his thigh and cooked a bowl of soup with his flesh for Chong'er. (Year 3)

(103) I sat *at the side of* the balcony, listening to the rain peacefully and enjoying the outside scenery. (Year 1)

(104) Qing Ming festival falls in early spring, on the 106th *day after the winter* solstice. (Year 3)

Description

Formulaic language with a description function is used to indicate quality, existence and situation in the texts. The majority of the descriptive formulaic strings and four-word clusters in this study are typically realised by a wide variety of prepositional and NPs (105) (106) (107).

(105) There is no doubt that the Dragon Boat Festival is *one of the most important* festival in China. (Year 1)

(106) It now has 4.5 million registered volunteers who have provided more than 4.5 *billion hours of volunteer* work. (Year 1)

(107) This film presents us *the coming of an era of* challenging and overthrowing traditional ethic. (Year 3)

Action

This study made some changes in terms of formulaic language with an ideational function. More specifically, there is a sub-category called "procedure" in previous studies to show how the research was carried out or to indicate events, actions and methods in the research. Taking the nature of the writing tasks used in this study, the researcher used term "action" instead of the original term "procedure" to indicate the action that taken out in the texts (Hyland, 2008a, b; Leedham,2011; Salazar, 2011).

Previous studies, for example, Salazar (2011), have found that the formulaic chunks serving the function of indicating events and actions mostly have past-tense and passive structures and the researcher considers this as a reflection of the fundamental concern of scientific research, which is giving an objective, unbiased and precise account of the experiment. However, in this study, on the contrary, the majority of the action strings and clusters are featured with active VPs (108) (109) (110). This is probably because of

the nature of the writing tasks, which are mostly based on the general knowledge or personal accounts and are written in an exposition way.

(108) And even now people *observe the festival* in memory of him. (Year 1)

(109) If everyone *turn a blind eye* to our trouble, and turn a deaf ear to our appeal for aid just because we are strangers to them. (Year 3)

(110) Travelling is a wonderful thing which can make one relax and *learn a lot of* things. (Year 1)

Quantification

The rest of the ideational formulaic strings and clusters serve the function of quantification, which means that they indicate the quantity (111), proportions (112), changes (113) or frequency (114) in the texts; they usually take the form of NPs, VPs or adverbial phrases.

(111) As a matter of fact, now *tens of millions of* people around the world have volunteered to help others in need. (Year 1)

(112) Helping others is the good thing and *most of* the people are surely proud of you. (Year 3)

(113) The story is full of unexpected twists and switchbacks, and opportunities for the audience to gear down and take a breath are *few and far between*. (Year 3)

(114) With the high speed development of technology, *a growing number of* people are much strongly inclined to a quick or disposable life. (Year 3)

6.1.2.4 Conversational function

There are two exceptional formulaic strings identified by the student writers in their written texts, which are actually used as politeness inquiries (115) and simple inquiries (116) and are mostly found in conversation rather than in the written register.

(115) *nice to meet you* (year 1)

(116) *what about my dreams?* (Year 1)

Both of these formulaic strings were used in the first writing task, in which students were asked to write a short piece in the first class of a new semester to introduce themselves. The genre of this type of writing is usually dialogical and colloquial rather than logical or formal.

This section has discussed the functional classification of two types of formulaic language identified in this study: formulaic strings and four-word clusters. It can be seen that student writers in this study made efforts to use formulaic language to perform different functions in their written texts, and some of these functions are realised differently from those in academic writing in the previous research (Hyland, 2008a, b; Leedham,2011; Salazar, 2011). For example, there are new sub-functions identified in this study, such as exemplification and generalisation. Furthermore, there are more active VPs with the present tense to describe events or actions, rather than depersonalised descriptions which most academic writing tasks expect. In the following sections, comparisons will be made between the two identification approaches, and year groups will shed further light on the formulaic language used in the student written texts. By this practice, Phase 4 of the analytical procedures will be realised and highlighted.

6.1.3 Comparison of the identification approaches: formulaic strings and four-word clusters

In this section, the students' written texts data is analysed and some tentative attempts are made to answer the first set of research questions, which focus on the features of formulaic language used by the students in their own English written texts.

Two types of data were derived from the collected written texts: the formulaic strings were identified based on the students' perception of what they thought was formulaic language in their own written texts, and the four-word clusters were extracted by a set of frequency-based and dispersion criteria. This section aims to discuss the two identification approaches employed in this study by comparing these two types of formulaic language, i.e. formulaic strings and four-word clusters, in order to illustrate the issues with two approaches (the speaker-internal and -external) to the area of formulaic language, which will be further discussed in the chapter 7. The comparison of the identification approaches includes the quantities, structural and functional and the distributions of formulaic strings and clusters.

6.1.3.1 Quantity

Table 6.19: The total number of formulaic strings identified in the Year1 and Year 3 corpora

Structure	**Formulaic strings Number (%)**		**Four-word clusters Number (%)**	
	Year 1	**Year 3**	**Year 1**	**Year 3**
Clausal	30 (14.29%)	13 (16.46%)	26 (16.64%)	36 (13.05%)
Phrasal	165 (78.57%)	187 (81.65%)	128 (81.01%)	234 (84.57%)
Other	15 (7.14%)	20 (9.09%)	4 (2.53%)	6 (2.18%)
Total	210 (100%)	220 (100%)	158 (100%)	276 (100%)

It can be seen from Table 6.19 that the total number of formulaic strings identified in the Year1 and Year 3 corpora are 430; whereas the total number of four-word clusters extracted by a corpus-based approach is 434. Therefore, the total amount of formulaic language (i.e. formulaic strings and four-word clusters) identified by those two approaches is very similar. Also, for the formulaic strings, the numbers remain very similar from Year 1 to Year 3, i.e. 210 and 220 respectively. It can be concluded that there is no substantial improvement or changes in the quantity of formulaic language used between the junior year (Year 1) and the senior year (Year 3) student English written texts.

To examine the differences in the nature of the two identification approaches, in this section, the focus will be on the two types of formulaic language identified by the students and the corpus linguistics approaches. In order to achieve compatibility between the two types of formulaic language, the researcher decided to use percentage and proportion in the identified strings and clusters to compare the results and findings, while in the next section, the numbers can be used to compare directly when it comes to the comparison across year groups (see section 6.1.4).

6.1.3.2 Structure

Figure 6.1 below shows the proportions of broad structural categories in formulaic strings and four-word clusters in the Year 1 and Year 3 corpora.

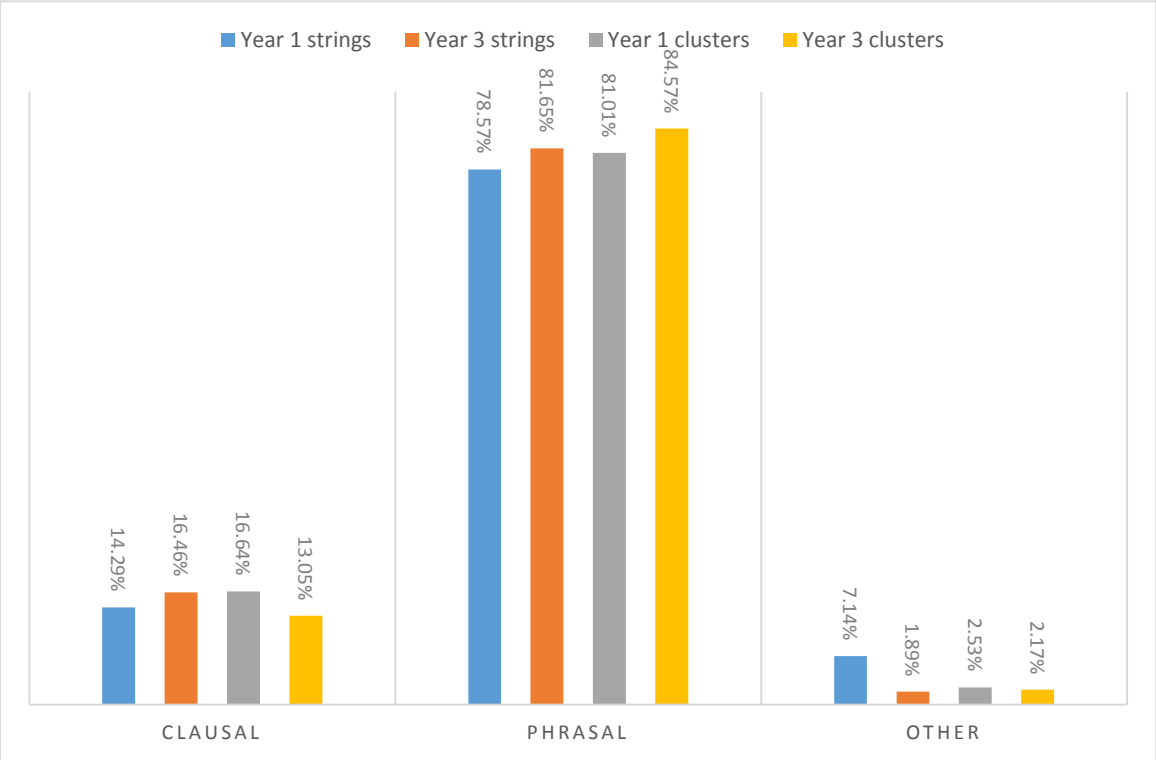

Figure 6.1: The proportions of broad structural categories in formulaic strings and four-word clusters in the two corpora

It is noted that Figure 6.1 above illustrated the general trend shared between the two identification approaches that the most identified strings and clusters are with phrasal structures, followed by those with clausal structures and the strings and clusters with other structures take up the smallest proportion of the whole. This finding indicates that formulaic language with phrasal structures constitutes a large section of the overall use of formulaic language by the students in this study.

Furthermore, as can be seen in Figure 6.1, there is no significance in the proportions of identified formulaic strings and four-word clusters with phrasal and clausal structures. Only in the category of "other", are there some differences. Next, further analysis of the sub-categories under each broad category will be presented and discussed.

Phrasal

Within the broad structure of phrasal based formulaic language, there are four sub-categories: VP-based, NP-based, PP-based, and Adj. P/Adv. P in the identified formulaic strings and four-word clusters in two corpora, as shown in Figure 6.2 below.

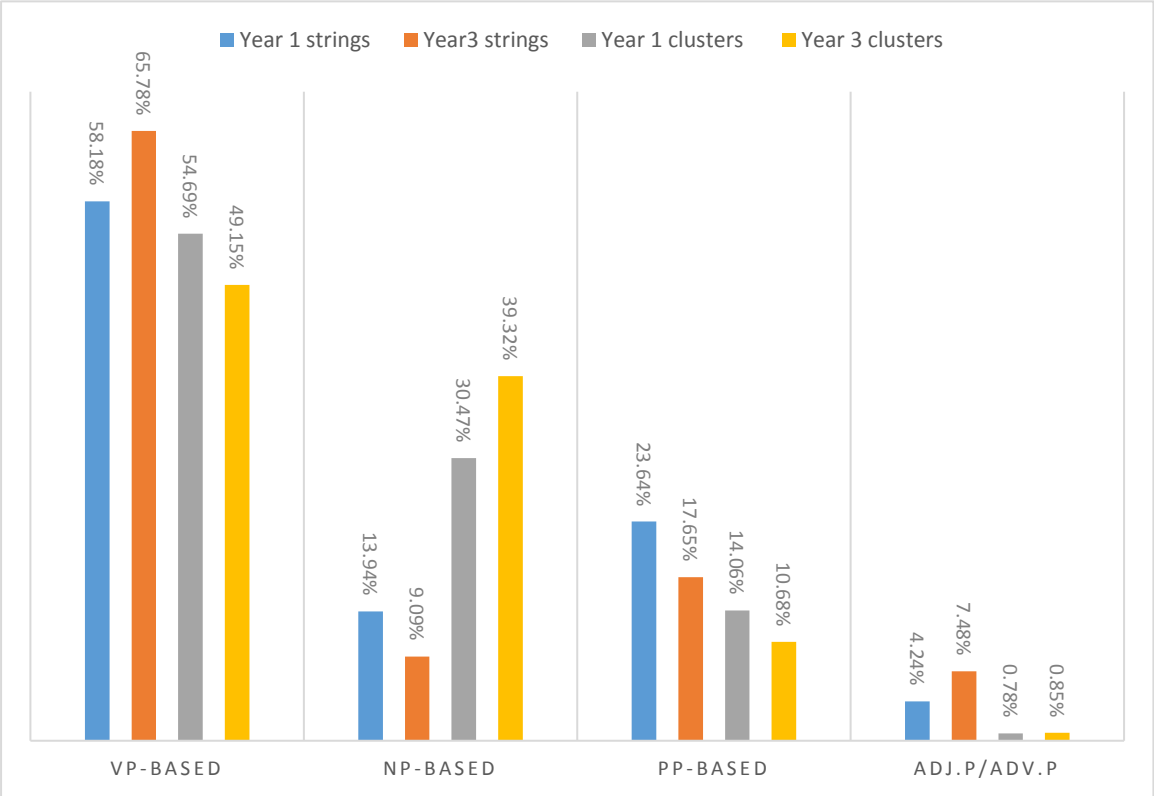

Figure 6.2: Proportions of the sub-categories under the phrasal structure

There are some features shared between the two identification approaches, as shown in Figure 6.2. For example, VP-based are the most identified phrasal structures in the formulaic strings and four-word clusters, they take up more than half of the whole amount of identified formulaic language with phrasal structures. Whereas Adj. P/Adv. P are the least identified phrasal structures in the formulaic strings (Year 1, 4.24% and Year 3, 7.48%) and clusters (Year 1, 0.78% and Year 3, 0.85%).

As for the NP-based and PP-based phrasal structures, the results show some divergence. In the formulaic strings, there are more PP-based structures (Year 1, 23.64% and Year 3, 17.65%) than NP-based (Year 1, 13.94% and Year 3, 9.09%) structures; whilst in the four-word clusters, there are more NP-based (Year1, 30.47% and Year 3, 39.32%) structures than PP-based (Year 1, 14.06% and Year 3, 10.68%) structures. The reason for this is

probably because of the extraction criteria for clusters, one of which is to stop at punctuation in Wordsmith 6.0, where most clusters with PP-based phrasal structures come before or after punctuation (also see Salazar, 2011).

Clausal

The proportions of the six sub-categories under the clausal structure are shown in Figure 6.3 below. These six sub-categories are to-clause, as-clause, if-clause, that-clause, WH-clause and 1st/2nd person pronoun +dependent clause. Among these clausal structures, the structure that-clause is not found in the formulaic strings.

Overall, the findings from the two identification approaches are mixed. For example, in the formulaic strings, the most identified structure is the WH-clause structure (Year1, 40% and Year 3, 46.16%), whereas that-clause is the least identified structure (no strings found with this structure). In the clusters, to-clause (Year 1, 46.15% and Year 3, 52.78%) and WH-clause (Year 38.46% and Year 3, 27.78%) are the most identified structures.

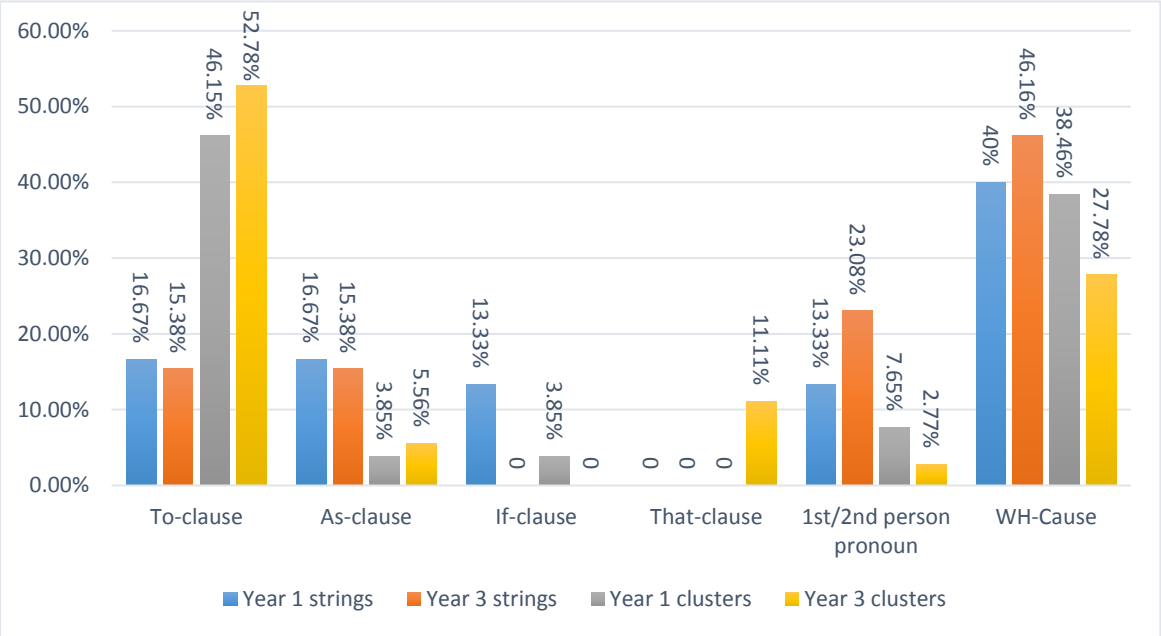

Figure 6.3: Proportions of the sub-categories under the clausal structure

Other structures

Findings regarding the third broad structure “other” from the two identification approaches are the most divergent compared with other two broad structures, as shown in Figure 6.4.

First, ill-formed clusters and formulae are not found in the clusters. The reason for this is probably that the ill-formed clusters are not frequent enough to be counted in the final list for analysis; as for the formulae, this is probably due to the fact that the length of clusters for this study is four words, while most formulae are complete sentences, usually with varied usage (see section 6.1.1.3). Thus, it also maybe because of the infrequency of occurrences in the corpora.

Second, in the formulaic strings, the final sub-category “others” only takes up a small portion of the whole broad “other” structure category (Year 1, 13.33% and Year 3, 15%), whereas in the clusters, the sub-category “others” takes up a large section of the whole broad structure (Year 1, 75% and Year 3, 50%). The possible explanation for this is the small size of both corpora, and the identified clusters seem to be more fragmental and meaningless than those findings from large corpora.

The final divergent findings are under the sub-category templates. The identified formulaic strings with this structure include both macro- and micro-level (see section 6.1.1.3), whereas in the four-word clusters, templates occur only at the micro-level.

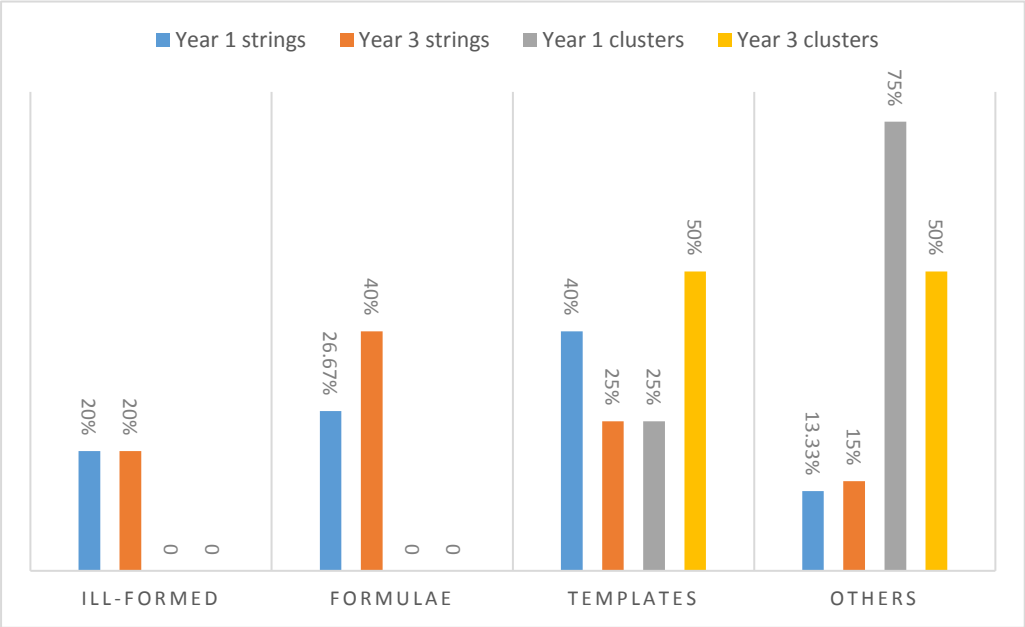

Figure 6.4: Proportions of the sub-categories under the clausal structure

6.1.3.3 Function

All the identified formulaic strings and four-word clusters perform three broad functions: textual, interpersonal and ideational, as discussed in section 6.1.2 and in appendices 12 and 13. Figure 6.5 below shows the proportions of broad functional categories of identified formulaic language, i.e. formulaic strings and four-word clusters in the two corpora (Year 1 and Year 3).

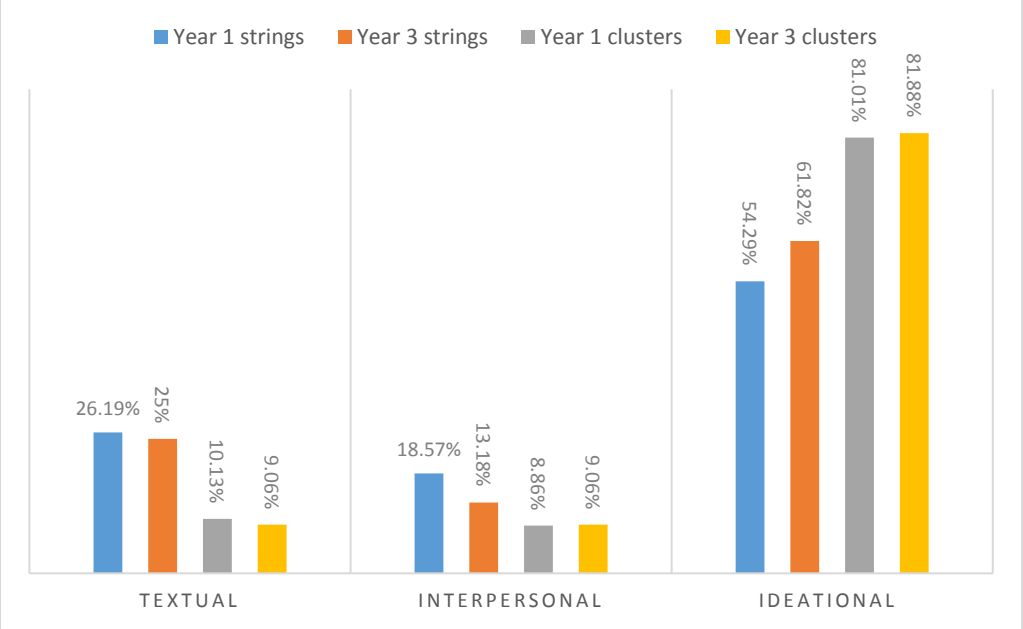

Figure 6.5: Proportions of broad functional categories of identified formulaic strings and four-word clusters in the Year 1 and Year 3 corpora

From Figure 6.5, it is clear that the majority of identified functions are ideational in both the formulaic strings (Year 1, 54.29% and Year 3, 61.28%) and the four-word clusters (Year 1, 81.01% and Year 3, 81.88%); followed by the textual function, where the strings take up 26.19% and 25% in the Year 1 and Year 3 corpus respectively, and the clusters take up 10.13% in the Year 1 corpus and 9.06% in the Year 3 corpus. The interpersonal function is the least common found one among the three broad functions in both formulaic strings (Year 1, 18.57% and Year 3, 13.18%) and four-word clusters (Year 1, 8.86% and Year 3, 9.06%).

Apart from the shared trend between two identification approaches above, there are still some divergent findings. For example, in ideational and interpersonal functions, identified strings take up more proportions than the clusters within the same function; whereas in the textual functions, clusters take up a greater proportion than the strings. Next, the sub-functions under each broad functional category will be discussed

Textual

Figure 6.6 shows the sub-functions under textual function and the proportions of identified formulaic strings and four-word clusters in each sub-functional category.

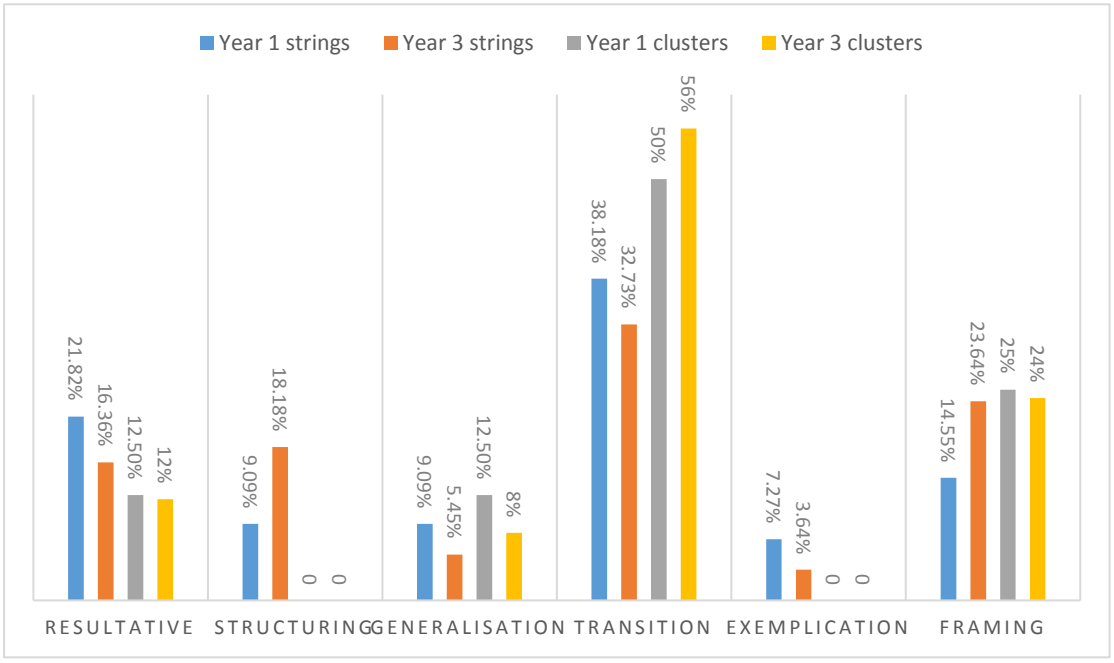

Figure 6.6: Proportions of formulaic strings and four-word clusters under textual function in the Year 1 and Year 3 corpora

One convergent finding from the two approaches is that the transition function plays an important role in the student written texts from both years, as the strings (Year 1, 38.18% and Year 3, 32.73%) and clusters (Year 1, 50% and Year 3, 56%) with the transition function took up the largest proportion of all formulaic language with the textual function. Also, generalisation (strings: Year 1, 9.09% and Year 3, 5.45%; clusters: Year 1, 12.5% and Year 3, 8%) is the least common found function in both the strings and clusters.

In addition, structuring and exemplification are not found in the extracted clusters. The possible explanation for this could be the extraction criterion "stop at punctuation" and the length of the clusters, as formulaic language with an exemplification function and structuring function is usually with PP near punctuation (see section 6.1.1 and 6.1.2).

Interpersonal

The interpersonal function was found to be the least common among the three broad functions. Overall, the proportions of identified formulaic strings and four-word clusters with the stance function are far greater than those with the engagement function. In addition, these proportions in each year group are similar, as shown in Figure 6.7 below.

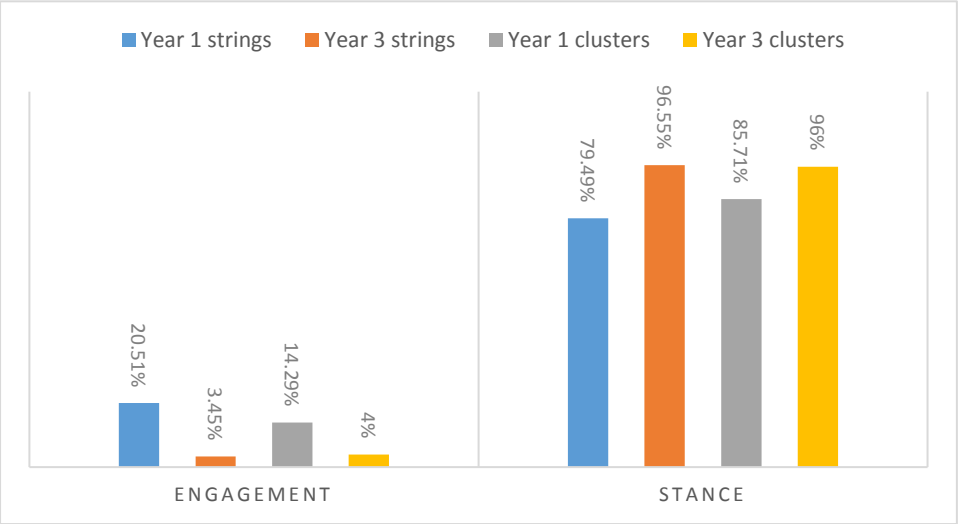

Figure 6.7: Proportions of formulaic strings and four-word clusters under interpersonal function in the Year 1 and Year 3 corpora

Ideational

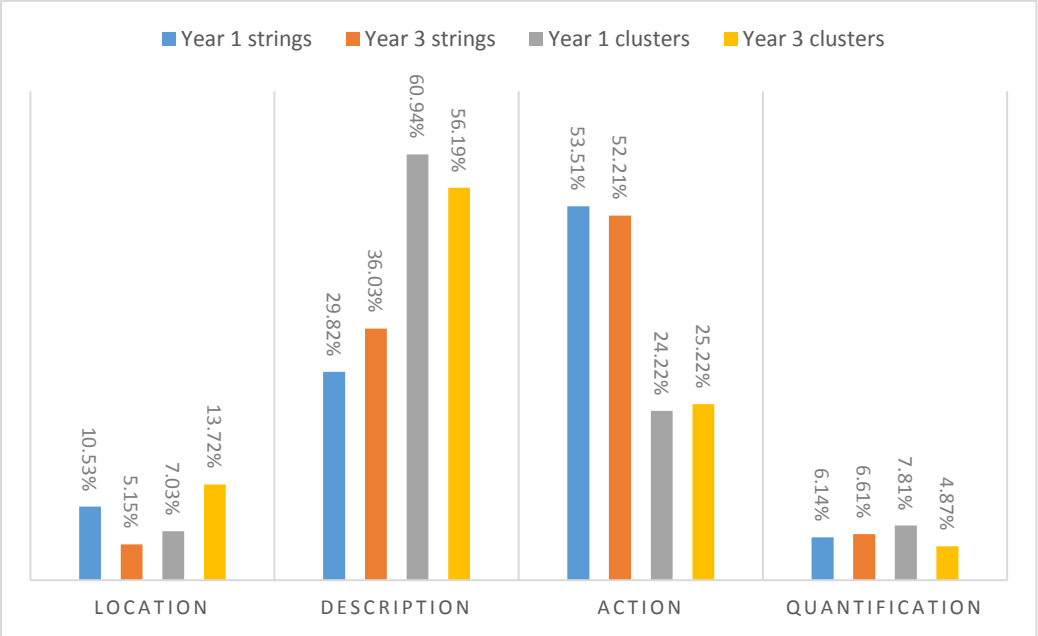

Figure 6.8: proportions of formulaic strings and four-word clusters under ideational function in the Year 1 and Year 3 corpora

The figure above presents the proportions of formulaic strings and four-word clusters under the ideational function in the two corpora and shows very divergent results. For instance, formulaic strings (Year1, 53.51% and Year 3, 52.21%) take up larger percentages than the four-word clusters (Year 1, 24.22% and Year 3, 25.22%) in the action function, whereas in the description function, the four word clusters (Year 1, 60.94% and Year 3, 56.19%) take up a greater proportion than the strings (Year 1, 29.82% and Year 3, 36.03%).

In this section, the comparison of the two types of formulaic language identified through the two identification approaches is realised through examining the percentage of identified formulaic strings and four-word clusters in terms of quantity, structural and functional categorisation. The results show some shared trends and at the same time some divergent findings.

6.1.4 Comparison of formulaic language use: Year 1 vs. Year 3

In this section, the comparison will be made based on the numbers rather than percentage or proportion. The reason for this is that it is possible to examine the

differences between year groups by checking the numbers of identified formulaic strings and four-word clusters directly. This practice can provide the evidence to answer research question 1 (d).

6.1.4.1 Quantity

Table 5.8 shows the total numbers of formulaic strings (Year 1, 210; Year 3, 220; total number, 430;) and four-word clusters (Year 1, 1588; Year 3, 276; total number, 434;) identified by two approaches in the Year 1 and Year 3 corpora respectively.

In Chapter 5, the average lengths of the texts in Year 1 and Year 3 have been worked out. The average length of a text in the Year 1 corpus is 56058/300=187 (186.87); whereas the average length of a text in the Year 3 corpus is 64186/251=256 (255.72). Then, the average of different types of formulaic language identified in each text are shown in the table below:

Table 6.20: The average of different types of formulaic language identified in per text

	Strings	**Clusters**
Year 1	(210/300=) 0.7	(158/300=) 0.75
Year 3	(220/251=) 0.87	(276/251=) 1.09

It seems that Year 3 students used more formulaic language than the Year 1 students, in terms of the numbers of identified formulaic strings and four-word clusters in the two corpora, as shown in Table 6.20. However, it is difficult to tell which group of students used more formulaic language. The reasons are as follows:

First, as discussed in section 5.8, the total numbers of identified formulaic strings and four-word clusters are similar, especially the numbers of formulaic strings, only 10 strings less in the Year 1 corpus than in the Year 3 corpus. However, the difference in the numbers of clusters extracted from the two corpora is large, i.e. 276-158 =118.

Second, as shown in Table 6.20 above the average numbers of identified formulaic strings and clusters are also similar, ranging from 0.7 to 1.09; furthermore, the results regarding the average amount of formulaic language in each text are without substantial differences: slightly more strings and clusters in the Year 1 corpus than those in the Year 3 corpus.

In addition to this, consider the difference sizes of the two corpora, it is necessary to normalise before the direct comparison. The normalised number of formulaic strings in the Year 3 corpus is 56058/64186* 220=192.06. The normalised number of four-word clusters in the Year 3 corpus is 56058/64186*276= 240.95.

These normalised results show that more strings identified in the Year 1 corpus than the Year 3 corpus, whereas fewer clusters identified in the Year 1 than the Year 3 corpus.

Next, the differences in the two year groups will be compared by structure, function, and distribution.

6.1.4.2 Structure

Overall

From Figure 6.9, it can be seen that the most frequently identified formulaic clusters and four-word clusters are those with phrasal structures, followed by those with clausal structures. The final structure “other” is the least identified.

In terms of the amount of formulaic language used in the written texts, the two groups, share some general trends. For instance, in the structural category phrasal and others, there are more formulaic strings and clusters identified in the Year 3 corpus (Phrasal: strings, 187, clusters, 234; Other: strings, 20, clusters, 6) than in the Year 1 corpus (Phrasal: strings, 165, clusters, 128; Other: strings, 15, clusters, 4). On the other hand, there are some divergent findings. For example, in the clausal structure category, there are twice the number of clusters in the Year 3 (36) corpus than in the Year 1 corpus (13); while more strings are identified in the Year 1 (30) corpus than in the Year 3 corpus (26). However, the total number of clusters and strings in the Year 1 corpus (20+26=56) and in the Year 3 corpus (12+36=47) are very different, with slightly more clusters considered as formulaic language.

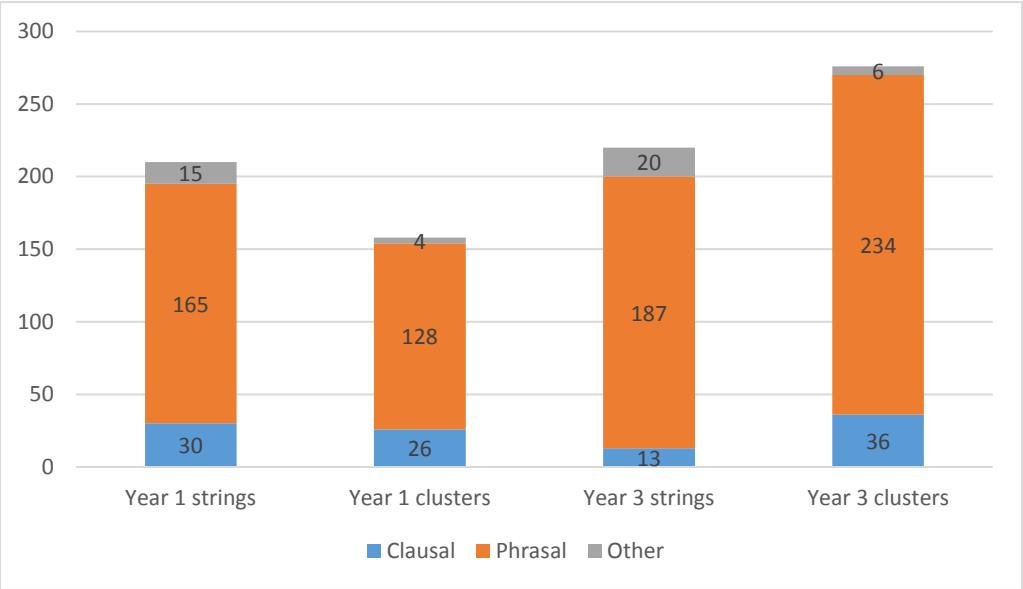

Figure 6.9: Numbers of formulaic strings and four-word clusters identified in the Year 1 and Year 3 corpora

Next, the differences across the year groups will be examined in each broad structural category.

Phrasal

Figure 6.10 below shows the numbers of identified formulaic strings and four-word clusters in the sub-category under the phrasal structural category.

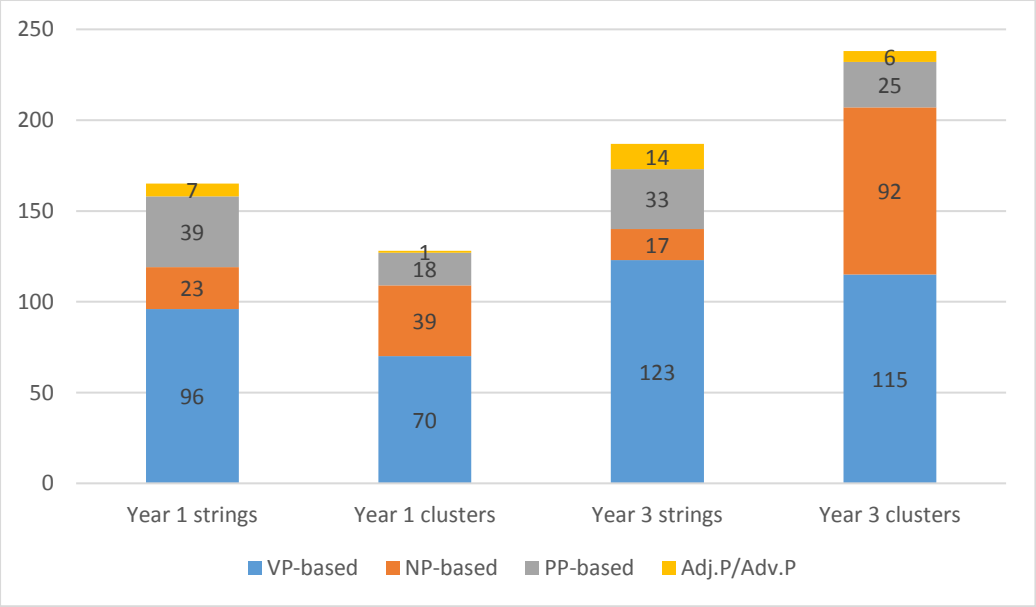

Figure 6.10: Numbers of identified formulaic language with phrasal structure in the Year 1 and Year 3 corpora

In Figure 6.10, it is clear that both groups shared some trends in using formulaic language with a phrasal structure: first, most identified formulaic language (i.e. formulaic strings and four-word clusters) have a VP-based structure. This corresponds to the results shown in Table 6.2. Also, it can be concluded from Tables 6.5 and 6.6 that the most popular structures within the VP-based category are active VPs in both corpora. On the contrary, the least identified phrasal structure is Adj. P /Adv. P in both year group corpora. Second, the Year 3 group tended to use more formulaic language with VP-based and Adj. P/Adv. P structures. This could be because the more experienced writers (Year 1 group students) tend to have a larger vocabulary and describe events or actions more precisely.

As for the divergent aspect, it is mainly in the NP-based and PP-based structures. In terms of NP- and PP-based formulaic strings, there are more identified in the Year 1 corpus (NP-based, 23; NP-based 39) than in the Year 3 corpus (NP-based, 17; PP-based 33). This finding contradicts previous studies (Biber, et al, 1999; Hyland, 2008a;), in which researchers claim that more proficient and/or experienced writers use more formulaic language with NP-and PP-based structures in their academic writing.

Clausal

As can be seen in Figure 6.11, the findings regarding formulaic language with clausal structures in the two corpora are quite divergent.

For the shared trend, the Year 1 group students tend to use more formulaic language with the WH-clause (strings, 12; clusters, 10) than the Year 3 group (strings, 6; clusters, 10). Similarly, the 1st/2nd person pronoun+ dependent clause structure was identified more in the Year 1 corpus (strings, 4; cluster, 2) than in the Year 3 corpus (strings,3; clusters, 1).

However, with regard to the to-clause and as-clause structures, the results are mixed. In addition to this, formulaic language with that-clause structures is not found in the formulaic strings, and if-clause structures are not found in the four-word clusters.

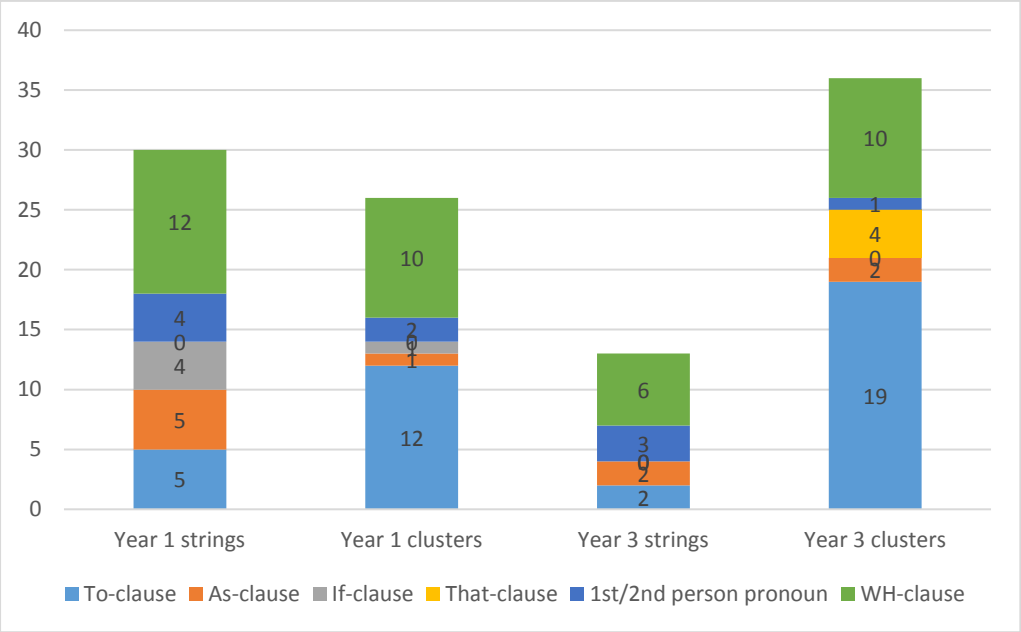

Figure 6.11: Number of formulaic language identified with clausal structures in the Year 1 and Year 3 corpora

Others structures

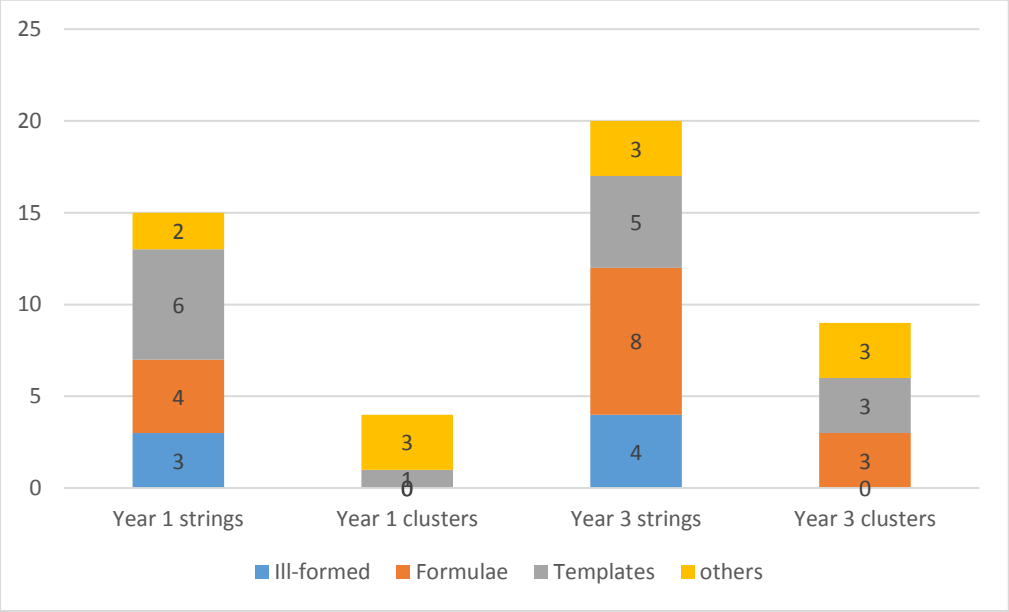

Figure 6.12: Numbers of formulaic language identified with other structure in the Year 1 and Year 3 corpora

According to Figure 6.12, overall, the Year 3 group students used more formulaic language than the Year 1 students, as two sub-structural categories are not found in the clusters in either corpus: ill-formed and formulae. The nature of the methodology

probably results in this The two groups found an equal amount of formulaic language with other structures. (also see section 6.1.3.2).

6.1.4.3 Function

This section will discuss the group differences in terms of function. It begins with a discussion of the overall trend regarding the three broad functional categories, then it takes further steps to examine the sub-structure categories under each broad function.

Overall

In general, as shown in Figure 6.13 below, it is clear that the Year 3 group use more formulaic language (including formulaic strings and four-word clusters) than the Year 1 group. Also, based on Figure 6.5, formulaic language with an ideational function in the Year 3 group is used 20% more than in the Year 1 group, whereas textual and interpersonal functions were used around 10% less in the Year 3 compared to those in the Year 1 group. It is not clear which factors cause these differences, in other words, factors like the nature of the writing task, the size of vocabulary and the expertise in the English writing are all possible contributors to this situation.

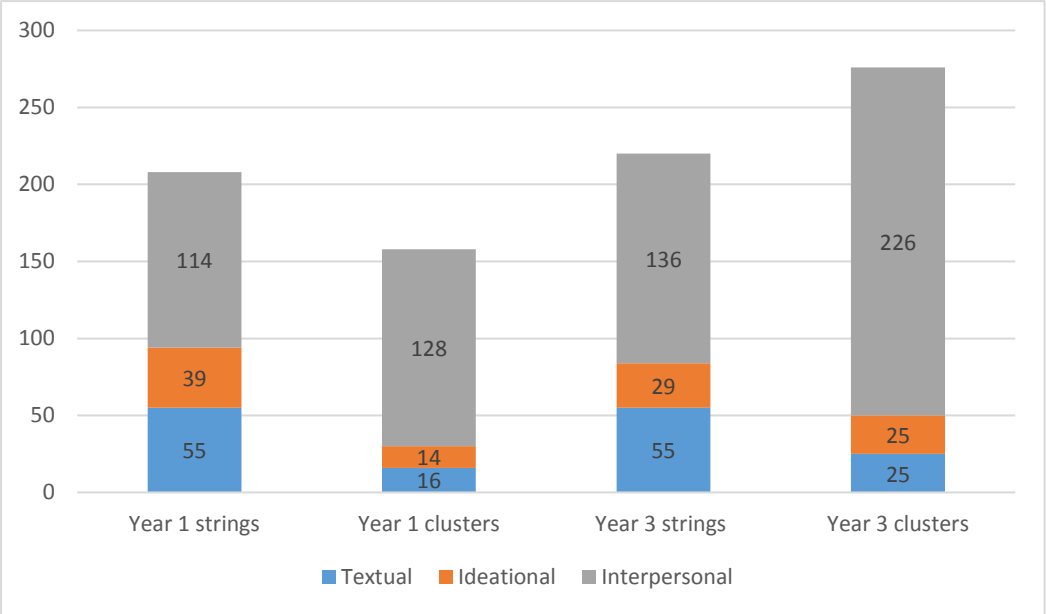

Figure 6.13: Numbers of formulaic language identified with three broad functional categories in the Year 1 and Year 3 corpora

Textual

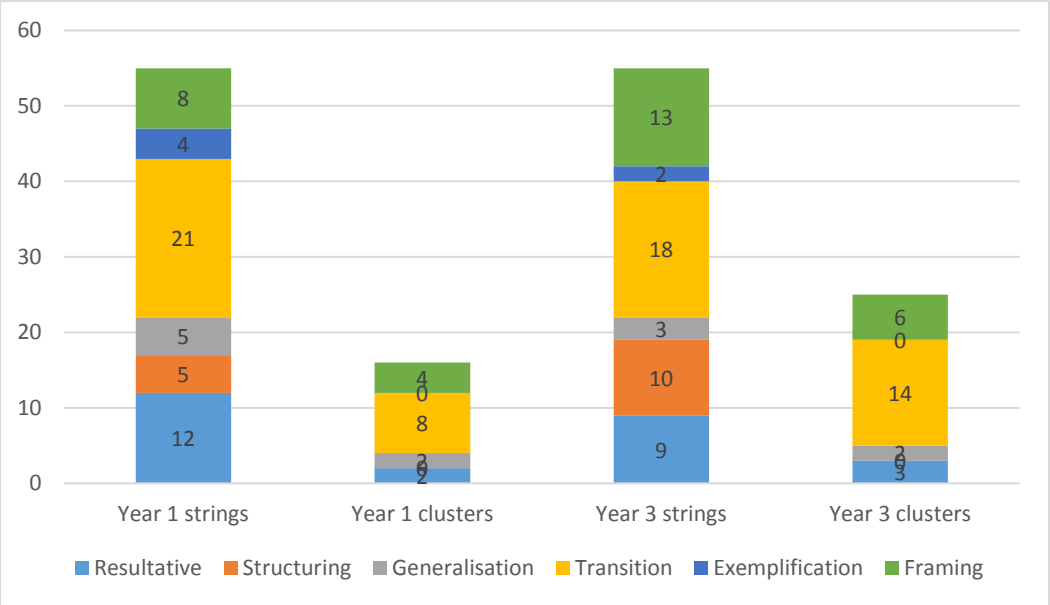

Figure 6.14: Numbers of formulaic language identified with textual functions in the Year 1 and Year 3 corpora

Figure 6.14 shows the numbers of formulaic strings and clusters in the two corpora. Overall, the largest amount of formulaic language used by both groups had a transition function. Roughly, the Year 1 group tended to use more formulaic language with structuring, generalisation exemplification functions than the Year 3 group. However, regarding the other three functions---- resultative, transition, and framing, the findings show a mixed picture among which, clusters with a structuring function are not identified (also see section 6.1.3.3).

Interpersonal

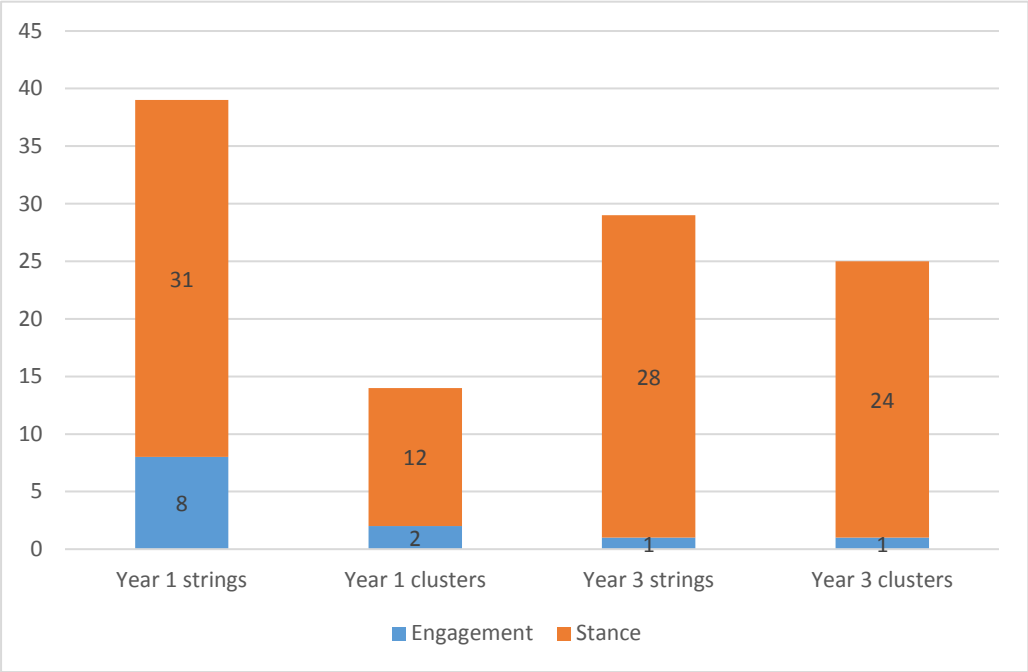

Figure 6.15: Numbers of formulaic language identified with interpersonal functions in the Year 1 and Year 3 corpora

As the least identified function in both corpora, the interpersonal function mainly indicates the writer's stance, including attitudes, evaluation, and judgement; as well as the engagement with the readers of texts. In Figure 6.15, it is noted that the majority of the formulaic language with an interpersonal function is used to express the writer's stance in this study. Compared to the Year 3 cohort, the Year 1 students tend to use more formulaic language with an engagement function.

Ideational

The ideational function is the most identified function among the three broad functions. In its four sub-structural categories, the description and action functions are the two most identified categories. Overall, the Year 3 students use more formulaic strings and four-word clusters with an ideational function than the Year 1 students. It can be assumed that the Year 3 students have a richer repertoire to express themselves or to refer to the real world in their texts.

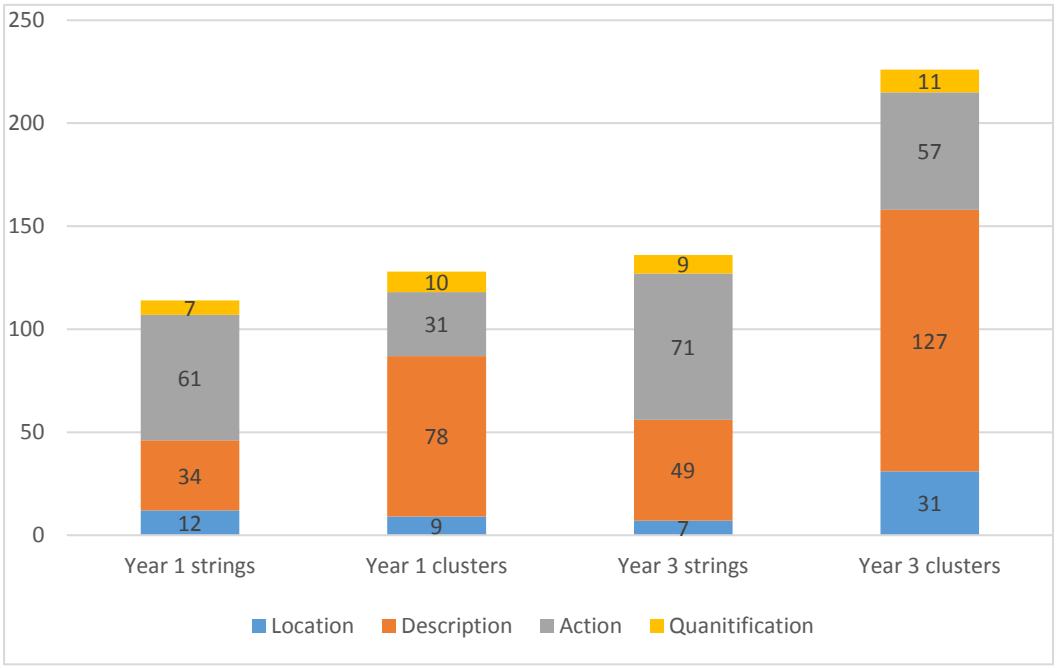

Figure 6.16: Numbers of formulaic language identified with ideational functions in the Year 1 and Year 3 corpora

6.1.5 Shared formulaic language: identified by both approaches and used in both corpora

Appendix 14 lists the core formulaic language identified by both approaches (student identified formulaic strings and four-word clusters extracted by the corpus linguistics approach) and the shared formulaic strings and clusters in both year groups. "Shared strings / clusters", and "core formulaic language" are the terms used to refer to different occasions of shared formulaic language.

One is shared formulaic strings and clusters. This very straightforwardly refers to the strings or clusters that are the same or overlapped in the two corpora. Whereas the term "core formulaic language" refers to the same or overlapping strings and clusters identified in one corpus (more discussion on overlap see section 5.8). The tables below illustrate the statistics:

Table 6.21: Numbers of shared formulaic language identified by both approaches in the structural categories

	Formulaic strings	**Four-word clusters**
Clausal	5	4
NP	2	6

Other	9	2
PP	18	8
VP	8	20
Total No.	42	40

Table 6.22: Numbers of shared formulaic language identified by both approaches in the functional categories

	Formulaic strings	**Four-word clusters**
Ideational	10	14
Interpersonal	4	6
Textual	28	20

Table 6.23: Numbers of core formulaic language in two corpora with structural categories

	Year 1	**Year 3**
Adj. P/Adv. P		1
Clausal	5	3
NP	6	7
Other	6	2
PP	12	8
VP	1	9
Total No.	30	30

Table 6.24: Numbers of core formulaic language and clusters in two corpora with functional categories

	Year 1	**Year 3**
Ideational	8	13
Interpersonal	10	3
Textual	12	14

As illustrated in Tables 6.21 and 6.22, it is clear that the total numbers of shared strings and clusters in both corpora are similar, as there are 21 pairs of strings and 20 pairs of clusters. From Tables 6.23 and 6.24 we can see that the total amount of formulaic language identified in the Year 1 and Year 3 corpora are the same----both of them have 15 pairs. This can indicate that core formulaic language remains very stable regardless of the change in identification approaches or the language proficiency level of the language users. Next, a closer look will be taken into the structural and functional aspects of this core formulaic language.

It is found that both formulaic strings and four-word clusters are possible with incomplete structures. Clusters are known with incomplete structures and usually stretch between two phrasal spans. For example, the structures of the strings *I think it is a* and I *those who are in need* are not complete or stretches across phrases, the fragmental parts in these two strings, 'I think [*it is a*]' and '*those* [*who are in need*]', remind the language users of the grammatical component of the following part in the sentence, usually in a sentence with complex structure. By this reminder, language users can realise the accuracy of the grammar. At the same time, it reduces the load for the language users by giving more time to the information processing rather than to the consideration of grammaticality during the communication. In this way, formulaic strings, as one type of formulaic language, function as a structural material and make the students' writing task easier because they work with those ready-made sets of reminders and it can increase the fluency in the writing (Coxhead and Byrd, 2007: 134; Hüttner, 2007).

However, the clusters do not have this reminder function, for example, in the "other" sub-structural category, which shows the fragments from the corpus extracts in the clusters (see section 6.1.1).

Another significant finding is that the function of this core formulaic language is relatively fixed, but the structures can be less stable. As can be seen in Appendix 14, all the functions assigned to the pairs in shared formulaic language remain the same, whereas in the structure, due to the different boundaries that are drawn by the language users and the corpus tool, their structural categories may change accordingly, for instance, *on the one hand* (cluster) is considered to have a PP-based structure, whereas the string *one the one hand…on the other hand*… is assigned to the "Other structures" category as a template structure. However, their function remains the same, as a transition device, to add extra information to the text.

In terms of a comparison of these core formulaic strings and clusters, they share the trend that the most identified function is the textual one. However, in terms of

structure, the most identified core formulaic strings are PP-based, whereas the clusters are VP-based.

As for the comparison of the year groups, formulaic language identified in the two corpora covers the following five categories: NP-, VP-, and PP- based, clausal and other. For Year 1, the clausal structure is most identified, for Year 3, it is the VP-based structure. For both corpora, the most identified core formulaic language has a textual function.

Based on these two comparisons, it can be concluded that the textual function plays a role in the core formulaic language, and VP-based, PP-based and clausal are the most commonly identified structures.

6.1.6 Distribution of the structural and functional categories

In the last section, the relationship between the structural and functional aspects of the core formulaic language was discussed. In order to obtain a comprehensive picture, this section will focus on the distributional relationship between the structure and function of formulaic strings and clusters.

The figures below show the distribution of five structural categories, i.e. NP-, VP-, and PP- based, clausal and other, and three broad functional categories. In this way, it is also possible to compare the findings and results from the current study with previous research.

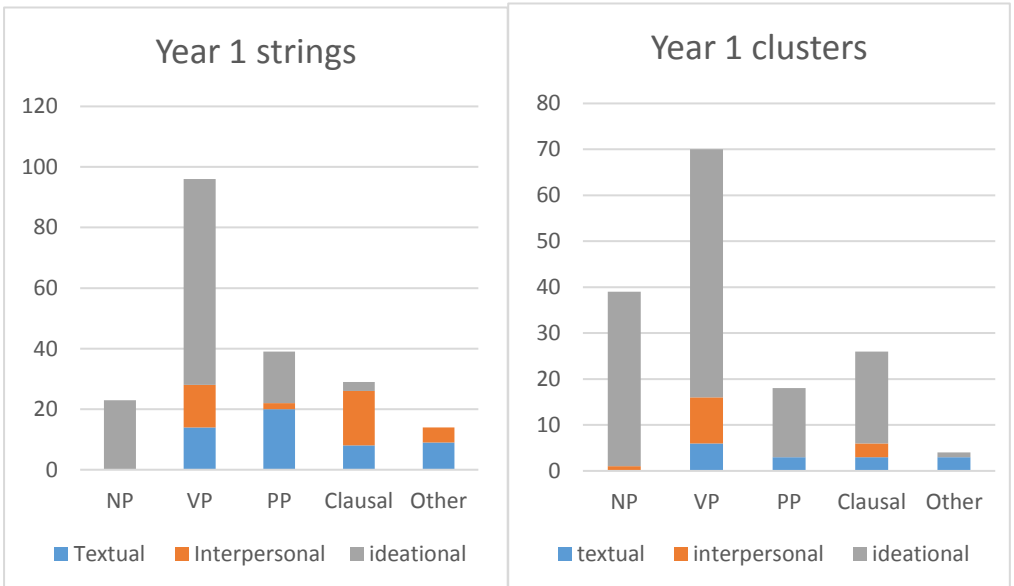

Figure 6.17: Distribution of structural categories across functional categories in Year 1 formulaic strings and four-word clusters

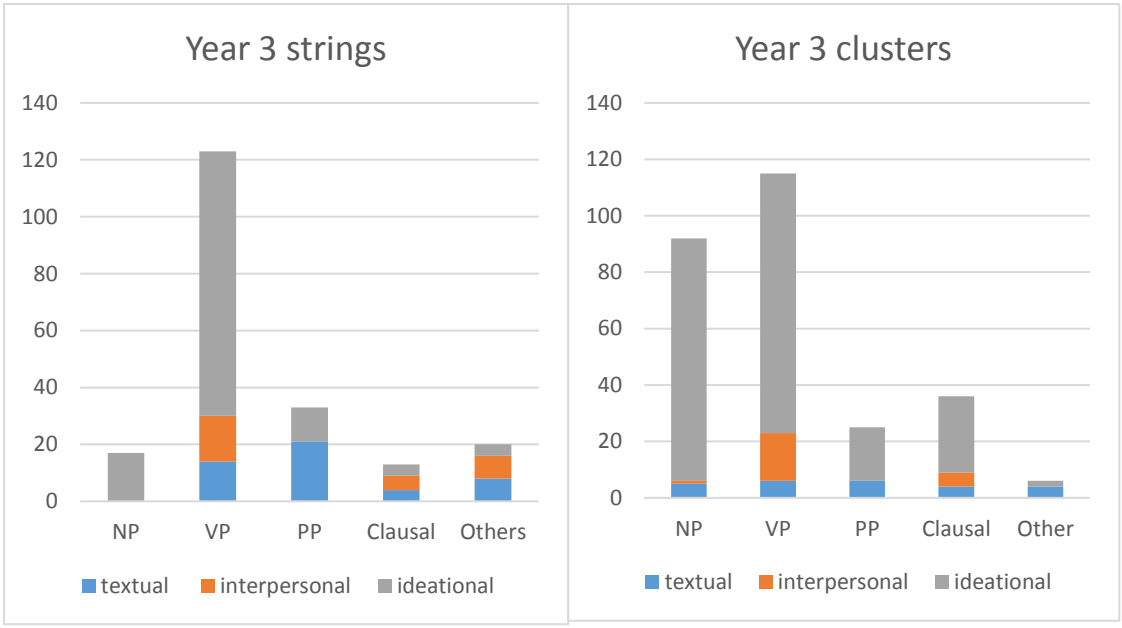

Figure 6.18: Distribution of structural categories across functional categories in Year 3 formulaic strings and four-word clusters

From these two figures, we can see that the identified formulaic language with VP-based and NP-structure mostly serves an ideational function; PP-based structures tend to serve a textual function or an ideational function; the clausal structure usually serves an interpersonal function, to establish the link between the reader, writer, and the text; formulaic language falls under “Other” and PP- based structural categories, which mostly serve a textual and an ideational function.

This section has presented the results based on the analytic procedure proposed in Chapter 5. First, it provides the structural and functional classifications for the student identified formulaic strings and four- word clusters extracted by the corpus linguistics approach respectively, with exemplars from the students’ written texts, which can provide a solution to research questions (1a) and (1b). It then turns to a comparison of the types of formulaic language, i.e. formulaic strings and four-word clusters, and the two corpora from two year groups. This practice can shed light on the answers to research questions (1c) and (1d). After this comparison, the researcher focuses on the shared formulaic language in the two identification approaches and two corpora respectively. Some features of core formulaic language are identified and elaborated.

Following this, the researcher also briefly discusses the distributional relationship between the structural

across functional categories. More discussion regarding the research questions will be further synthesised and provided in Chapter 7.

6.2 Interview findings: students' perceptions of formulaic language and its learning and teaching

This section examines and presents the results of the two rounds of interviews carried out with each of the 12 interviewees. The rationale for using semi-structured interviews in this mix-methods research has been discussed in the research methodology chapter (see section 5.6.2). In order to support and compare the results from the analysis of formulaic language in students' written texts, the interviews aim to examine their understanding of the concept of formulaic language, their formulaic language learning and use, the teaching of formulaic language that they have experienced in the English language classroom, as well as some considerations regarding the context of the research site.

6.2.1 Understanding of formulaic language

This section is comprised of two codes, the first code concerns the general understanding of formulaic language as a language phenomenon, which emerged from the first round of interviews where the interviewees were invited to make comments on formulaic language from three aspects: form, function, and examples when necessary; the second code concerns the six specific types of formulaic language (i.e. writing templates, idioms and fixed expressions, phrases, phatic expressions, and terminologies) that appeared in the second round of interviews, where the interviewees were expected to provide more information on each type of formulaic language.

The reason why I further examined the six types of formulaic language was as follows. Based on the literature review (see Chapter 2) I employed an inclusive view on formulaic language. Thereby, I drew on the study of Nattinger and DeCarrico (1992), who divided

formulaic language into four categories (poly words, institutionalised expressions, phrasal constraints, sentence builders) without clear-cut distinctions between each type. I also used Biber's and Hyland's work on lexical bundles (see section 2.3), and examined formulaic language at a macro-level, for example, writing templates, which help to structure whole texts; as well as at the micro-level, for example, terminologies and idioms and fixed expressions.

In the first round of interviews, words like fixed phrases, collocations, idioms and fixed expressions, proverbs, greetings, and templates were mentioned by the interviewees. Further, I divided these terms into six types of formulaic language according to the students' understanding.

Firstly, phrases, refer to a group of words, which tend to be together and form part of a sentence; they can form a constituent or function as a single unit in the syntax of a sentence, or can be collocations, including six sub-categories (adjective+ noun, noun + noun, verb+ noun, adverb + noun, verb+ prepositions, verb + adverb) for example, *crystal clear*. Secondly, idioms and fixed expressions, are mainly the combinations of words which have a figurative meaning because of common usage, which is different from their literal meaning. For example, spill the beans. Also, they can be used conventionally, such as proverbs or famous quotes. The third group comprises compound words, which are the product of word formation that creates compound lexemes, words like *breadknife*. Fourthly, writing templates can refer to the schematic structure or the outline of five-paragraph essays or specific sentences in a paragraph with open slots, as illustrated in Figure 6.21.

In recent years there have been many reports of ____. It turns a new chapter of ____ in China, and will have far-reaching effects in the forthcoming years. The biggest benefit, in my eyes, is that ____. In addition, ____. Finally, ____. Apart from the benefits mentioned above, we should also face several unavoidable challenges. In the first place, ____. In the second place, ____. What's more, ____. In summary, we should ____.

Figure 6.19: An example of a writing template

The fifth group comprises phatic expressions, which mainly refer to those expressions that are used in a fixed way and that are common among people or on one person in daily speech, including greetings, catch-phrases or some small talk in daily speech. For example, the expression "*how are you doing*" could serve a phatic function, probably performing a social task rather than conveying information in the conversation.

The final group includes terminology, which refers to the study of terms and their use. Terms are words and compound words that in specific contexts are given specific meanings—these may deviate from the meanings the same words have in other contexts and in everyday language, for example, Incoterms stands for international commercial terms. The reason why I added this in the interview is because the university has a technology and the foreign language department and the teaching and research objectives are focused on English for automotive trading and business English, which involves specific English content-based courses on business English and technical words relating to vehicles. In fact, business English and English for automotive technology dominate English language teaching and learning in that university, especially in the senior school years, Year 3 and Year 4.

Next, the analysis of two codes in the interviews will be presented. It is hoped that this analysis will contribute to the second set of research questions, which aims to find out the students' perceptions of formulaic language.

6.2.1.1 Understanding formulaic language in general

As previously discussed in section 2.3, there are two aspects of textual features that can help us to understand and define formulaic language: form and function. According to the transcriptions from the interviews, it is noted that all 12 interviewees talked about their understanding of formulaic language, especially in terms of form and function.

In terms of the form of formulaic language, interviewees from both groups mentioned fixedness as a feature of formulaic language (example 1); and they gave some examples of writing templates as formulaic language, which was important to them as they were

used in the written texts (example 2); in addition, interviewees generally considered "*the good sentence and phrases*" as formulaic language (INV 005 and INV 010 好词好句).

Example 1

1. "*Fixed collocations, fixed phrases, and fixed sentence structures* (INV 003)".[10]
2. "*Since he emphasised a lot about the fixed phrases, it might be formulaic language… fixed phrases, fixed sentence structure, and some famous quotes or proverbs* (INV 004)".

Example 2

1. "*In general, the first paragraph will present the main point, take the writing "the part-time job" as an example. I will present my viewpoint towards the part time job first, and then in the second paragraph, I will talk about opinions from other people. After this, I will focus on my own point. Finally, in the fourth paragraph, it is for the conclusion* (INV 004)".
2. "*Usually in the exam, the time is limited, so I do not write down any outline. I know I have to write the three-paragraph composition for the argumentative essay* (INV 007)".

Only two interviewees (INV 002 and INV 003) claimed that the high frequency was a feature of formulaic language, and they expressed this in example 3; whereas no one from the Year 3 group mentioned the frequency of formulaic language in use. From these statements, it is assumed that the interviewees demand the grammatical correctness, which is required in English exams all the time.

Example 3

1. "*For example, a collocation or sentence structure, appear again and again* (INV 002)".
2. "*Some collocations with high frequency in use* (INV 003)".

[10] The translated interview transcripts are in italic; the selected examples and the translation can be found in Appendix 12.

The Year 3 interviewees had a broader understanding of formulaic language and expanded the vision of formulaic language to the proverbs with some examples (example 4), as well as mentioning slang and phatic expressions such as greetings as possible forms of formulaic language (example 5); they also started to think about some formulaic language with an opaque meaning (example 6), which could be an obstacle to comprehension (see Chapter 2). In addition, native norms were not mentioned by the interviewees from the Year 3 group at all.

Example 4

1. "*Some Chinese proverbs, for example, no pain, no gain. Other proverbs, like I can use in this writing, the fragrance always stays in the hand that gives the rose* (INV 007)".
2. "*When it comes to the proverbs, I think the most frequent one we have come across is where there is a will, there is a way* (INV 010)".

Example 5

1. "*Fixed expressions, for example, be made up of, be composed of, and some others like the conjunctional words or phrases, like on the one hand…, on the other hand…; also, slang is formulaic language. I forget any example for that. I know less slang* (INV 009)".
2. "*The greetings are formulaic language, which you used to communication in speaking every day* (INV 010)".

Example 6

> "*Some collocations consist of several words, and you never know its meaning until you put that in the text, within the context of the language use, you can read and understand the content back and forth, and you can get the meaning as it is in the text* (INV 011)".

As for the function of formulaic language, interviewees showed their views on both sides. With regard to the positive function, firstly, more interviewees from the Year 1 group than the Year 3 group regarded formulaic language as a yardstick that L2 learners

had to follow since it could guarantee language use within the requirements (example 7). Second, it was convenient and handy to use the ready-to-use formulaic language, as it could be easily and quickly retrieved from the memory, which saved time and effort by avoiding the construction of new language (example 8). Third, six interviewees (INV 001, INV 003, INV 004, INV 008, INV 009, INV 010) from both groups contended that formulaic language could be a great addition to learners' language since it can increase correctness in terms of grammar (example 9). In addition, some interviewees claimed that sometimes, formulaic language can be used to produce funny or humorous effects in language use. This view is expressed in example 10.

Example 7

1. "*The writing templates could help structure my writing correctly and in accordance with the requirements in the writing tasks and guarantee the better score* (IVN 003)".
2. "*The writing templates help confine my language used in English writing within or following the requirement and pattern that the markers preferred* (INV 005)".

Example 8

1. "*If you can remember a template for English writing, and you can use some sentences in this templates directly, it is very handy, convenient, and save your time* (INV 004)".
2. "*When you write in English, you can remember these fixed items. When you need them, they just jump out in your mind, and you can use that in your writing. In this way, you can save time* (INV 005)".

Example 9

1. "It makes your language more vivid, and more native-like (INV 001)".
2. "*I think formulaic language is a great addition to my English. Suppose there are two compositions: One is without using any formulaic language, just a random writing, quite plain. The other one is with beautiful phrases, and formulaic language is used in a creative way in the text. It looks coherent and cohesive.*

Thus, the second compositions will have the higher mark and it makes the reader feel better (INV 003)".

3. "*In the TEM-4 exam, for example, if you use formulaic language frequently in different ways, it will be a great addition to your language and you mark will be increased. That's direct* (INV 009)".

Example 10

1. "In English, they have very funny phenomena, when you use idiom (INV 001)".
2. "Since lots of funny things could be expressed in a simple way (INV 001)".
3. "*Formulaic language is funny, fragmental and figurative* (INV 011)".

However, as for the downside of formulaic language, the interviewees also noted that the language was used in a generic and rigid way, as illustrated in example 11.

Example 11

1. "*It confines my language with lots of standards* (INV 002)",
2. *"I do not think it is good. There is too much limitation, just like writing a "Ba Gu Wen[11]", for example, this writing task. Also, I read a lot sample writing from the TEM-4, they all focused on the structures. Usually, when the teachers mark our English writing, they also check if you use language such as "first of all…second….third…"in your writing. I do not like this way of writing, as it is too rigid. I wish I can change it* (INV 008)".

Apart from the statement above, one interviewee from the Year 3 group commented that formulaic language had changed from a crutch to an additional language resource (see example 12).

Example 12

[11] Ba Gu Wen (eight-legged essay) is probably the most widely discussed traditional Chinese rhetoric. Starting from the ancient Ming Dynasty (1368-1644), the Ba Gu Wen (eight-legged essay) was the required form of essay writing in the imperial civil service examination. It is notorious for its rule-like format requirement, namely eight parts in the essay.

"*I feel that when we learned a new thing at first place, it was useful, anyway, we were not familiar with it. Yet, later on, I find that after you made some progress, you have to try to throw away this crutch finally. As they are the fixed templates, it is not good used by too many people* (INV 010)".

6.2.1.2 Understanding towards six types of formulaic language

In the last section, an analysis of students' understanding of formulaic language, in general, was presented, and it was noted that there were some divergences between the interviewees from the two groups in their understanding of the notion of formulaic language with regard to form and function. In this section, the analysis focuses on the six specific types of formulaic language (writing templates, idioms and fixed expressions, phrases, phatic expressions, compound words, terminologies) and their role in language learning.

As reported by the interviewees from both groups, writing templates can refer to the outlines of three- or five-paragraph essays or to specific sentences of a paragraph with open slots (example 2). In this category of formulaic language, all six interviewees from the Year 1 group provided some information about their understanding of writing templates (example 13).

Example 13

1. "*The writing templates, I think it refers to a basic frame, or the way of thinking, it also includes some sentences with a certain structure* (INV 002)".
2. "*In my opinion, in English writing, I usually use the writing templates, they generally are fixed. You then use those templates in your writing. You can change some parts in the templates. Thus I think it is the most importance one for me* (INV 006)".

As for the function of the writing template, two out of six interviewees (INV 002 and INV 006) from the Year 1 group, stated that to achieve a high score in exams, it is necessary to rely on writing templates (example 14).

Example 14

1. "*The writing templates, I think it refers to a basic frame, or the way of thinking, it also includes some sentences with a certain structure. All of these enable you to get higher marks in the writing* (INV 002)".
2. "*If you want to survive the exams, you have to remember phrases, idioms and writing templates very well. If you just follow your own ideas, you do whatever you want, you would not get a good score* (INV 006)".

As for the Year 3 group, the interviewees focused on the function rather than the form of writing templates, stating that it is very important for a written text, as it makes the writing clear, and is a great addition to the language (example 15).

Example 15

1. "*They can help you categorise your whole writing and present your ideas* (INV 009)".
2. "*They can make your writing clear to the reader* (INV 007)".

None of the interviewees from either group described idioms or fixed expressions specifically, however, they talked about their understanding of idioms and fixed phrases. According to the statements from the interviewees, idioms, and fixed expressions are usually fixed, with cultural connotations and remain within sentence-length (example 16).

Example 16

1. "*Idioms and fixed expressions are language and cultural specific, and contain rich connotation within them* (INV 003)".
2. "*There are couples of idioms and fixed expressions or phrases in one sentence. They can give you a sense of the structure in a sentence* (INV 009)".

With regard to the function of idioms or fixed expressions, the interviewees generally agreed on their positive role in language use: they can be a great addition to language and contribute to fluency in language production. They also play a role in communication with people from English speaking cultural backgrounds (example 17).

Example 17

1. "*People in the context of ESL used more idioms and fixed expressions than people in the context of EFL, like her, and she used them mostly in her writing rather than speaking* (INV 003)".
2. "*I would use them in writing, since the larger repertoire you have in idioms and fixed expressions, the more vividness you can get in your language and the more fluency you can achieve when you are writing* (INV 010)".

There was not much about the form of the phrases in the comments from the interviewees. Only two interviewees (INV 001 and INV 004) stated that phrases are the basic construction of a language (example 18). Apart from this, INV 004 provided his own opinion on what was not a phrase. S/He thought that common structures, like verb +noun, were not so called phrases (see example 18.2).

Example 18

1. "The phrases are the most basic ones. Since you learn the language to communicate with others, while, as a phrase, phrases that if you want to write something, if you want to speak, you cannot leave the phrases (INV 001)".
2. "*They are basic parts for constructing a sentence … for example, the combination of verb and noun. I don't think it is a phrase. For instance, the expression, play table tennis, play is a verb, table tennis is a noun phrase. For me, I don't think the expression play table tennis can be called a phrase. However, the teacher asked us to mark it and remember it as a phrase. For me, I think if you come across some expressions in your reading, you cannot say it is a phrase*" (INV 004).

As for the function of phrases, interviewees from both groups gave their own comments. As the Year 1 interviewees expressed it in example 19, phrases are conventionalised use of a language, and they are very practical in meaning making. The Year 3 interviewees noted that using phrases makes language creative and can impress the reader, as shown in example 20.

Example 19

1. “*One interviewee reported that phrases are useful when you express meanings* (INV 005)”.
2. “*You have to write down the right form, otherwise, if you write down phrases without the rules, you cannot get scores* (INV 006)”.

Example 20

1. “*Using some phrases are not so familiar to their peers, but easy to be understood by audience/readers, which probably could draw attention from the marker or teacher, since the student wants to be different from others in the class, and show his/her language use is ingenious* (INV 007)”.
2. “*They [classic phrases] just simplify the complex expressions, and people know it as it is popular, which makes your meaning expressed in a straightforward and impressive way* (INV 012)”.

From the example above, it is clear that the interviewees from Year 1 paid more attention to the translation process between two languages in terms of the function of phrases, whereas for Year 3 interviewees thought more about the use of phrases as an innovative way to improve language or make it more readable. Also, interviewees from both year groups mentioned the importance of using the correct form of phrases in the exams.

In terms of phatic expressions, none of the interviewees gave any detailed explanations of the form and function dichotomy. Instead, they demonstrated some rough understanding of phatic expressions. For the Year 1 group, the interviewees noted that phatic expressions were colloquial, and also that these expressions were not supposed to be used in writing, since it made the language informal and not appropriate, and English writing usually required formal language (example 21).

Example 21

1. “*Usually it is related to the impression that phatic expressions are not formal* (INV 002)”.

2. "*Not appropriate used in writing, since writing is a formal way to express meaning* (INV 004)".

The interviewees from Year 3 had greater knowledge of phatic expressions and explained that they were used mostly in spoken language and in communication with foreigners. They noted that phatic expressions in Chinese are different from those in English (example 22).

Example 22

1. "*Phatic expressions are not practical at all, since she did not have many opportunities to communicate with foreigners, thus she barely used them in a real conversation* (INV 011)".
2. "*I feel I haven't adapted myself in certain kinds of situation. For example, if some foreigners ask me "how are you", I cannot really respond him immediately since I was thinking of the answer to "how am I", so there is definitely cultural difference. It is very strange if you think phatic expressions like how are you in Chinese. In Chinese, we prefer asked others like, "how is everything going recently?" "Have you had your meal?" "Where are you going?" Such kinds of questions. If I was asked suddenly, I feel weird. I think because I don't get used to thinking in English and that's it* (INV 011)".

Compound words, according to the interviewees from the Year 1 group, are the product of word formation that creates compound lexemes, usually with transparent meanings (example 23).

Example 23

1. "*Compound is like the derivation, just a way of word formation* (INV 004)".
2. "You can see these compound words. First, you can see the meaning of this word, because if you know the two of the words, which is consist of it (INV 001)".

Among the interviewees from Year 3, some reported that the use of compound words would be a great addition to language production, as people do not usually use them that much; and they are a means of enlarging vocabulary size (example 24).

Example 24

1. "*The compound words, for example, some very new ones, if you can use compound words in a novel way, which make your language polished and special* (INV 007)".
2. "*Compound words can enrich and enlarge your vocabulary quickly and efficiently* (INV 012)".

As the last type of formulaic language that will be examined, terminologies was added by the researcher. Terminologies are words and compound words that in specific contexts are given specific meanings—these may deviate from the meanings the same words have in other contexts and in everyday language, for example, Incoterms stands for international commercial terms.

For the Year 1 group interviewees, four out of the six (INV 001, INV 003, INV 004 and INV 005) talked about their understanding of terminology. They did not give any examples, instead they talked more about the significance of terminologies used for different purposes. The statements on compound words are presented in example 25.

Example 25

> "*Terminologies, as each field has its own terminologies, such as science, social science, they have their subject-related terminologies. If you learn these terminologies, you will know the vocabulary of the whole area* (INV 004)".

Thus, the interviewees from this group contended that terminologies are mostly related to a specific subject area, and for the interviewees, learning them is a necessary introductory procedure to gaining knowledge in that area. The interviewees from the Year 3 group expressed their views in example 26.

Example 26

1. "*Terminologies involve certain knowledge of the specific area, for example, international trading, that is quite a fixed way of saying things* (INV 004)".

2. "*Terminology reduces the difficulties in the communication for both speaker and hearer, and makes all the interlocutors easier to be understood, especially within specific areas* (INV 007)".

Interviewees from this group also realised that terminologies are specialised knowledge based items, and require some special knowledge and accuracy in form. In addition, the Year 3 interviewees claimed that good command of terminologies in the area of study can improve efficiency in communication with people from specific areas.

In this section, the data regarding the understanding of formulaic language in two rounds of interviews were analysed and presented. In general, for the form of formulaic language, interviewees from both groups reported fixedness as a more important feature of formulaic language than frequency, and they also stated that writing templates are a type of formulaic language, as they are important in English writing. Also, the interviewees from the Year 3 group talked more broadly about the possible forms of formulaic language, such as greetings, proverbs, and famous quotes. In terms of function, interviewees showed their views on both sides. With regard to the positive role of formulaic language, the interviewees regarded formulaic language as a yardstick that L2 learners needed to follow since it could guarantee language use within the requirements. Also, it was convenient and handy to use the ready-to-use formulaic language, as it could be easily and quickly retrieved from the memory, which saved time and effort by avoiding the construction of new language. In addition, they contended that formulaic language could be a great addition to learners' language since it can increase grammatical correctness. In addition, some interviewees claimed that sometimes, formulaic language can be used for producing funny or humorous effects. One interviewee also provided a description of the changing role of formulaic language during the language proficiency developmental process.

As for the understanding of specific types of formulaic language, the interviewees showed that they were familiar with writing templates, and talked about them in terms of form and function. However, the interviewees were only able to give a rough description of the rest five types of formulaic language.

6.2.2 Formulaic language learning

In this section, the data regarding general learning aspects and the learning of the six types of formulaic language will be presented. In the analysis of the general learning of formulaic language, four themes emerged from the first round interviews: learning process, strategies and sources used for learning and the interviewees' comments on the importance of formulaic language; whereas in terms of learning specific types of formulaic language, the learning preferences were elicited in the second round of interviews and then were analysed in SPSS. Also, in the interviews, the interviewees were encouraged to talk about the strategies used in the six specific types of formulaic language.

According to the information provided by the interviewees, there were four themes extracted from the transcripts and coded under "formulaic language learning": the role of formulaic language in language learning, formulaic language learning processes, formulaic language learning strategies, and learning sources used in learning.

6.2.2.1 Is formulaic language important?

In this section, I will present the answers from the interviewees to the question "Is formulaic language important in language learning or not?". As in the interview, I asked all the interviewees a general question: "Do you think that the listed formulaic language categories are important to your language learning?".

All 12 interviewees gave me their answers. All 6 interviewees from Year 3 confirmed that the listed formulaic language was important to their English language learning. The interviewees from Year 1 provided different answers. One of the six said that most of them were important, and two of the six were not aware of formulaic language until they took part in this research.

As it shows, not all of the interviewees thought that formulaic language was important for English language learning, nor did all the interviewees have an awareness of formulaic language from their learning of English. In other words, there was no

consensus of student opinion on the role that formulaic language plays in English language learning.

6.2.2.2 Formulaic language learning process

As for the general learning process of formulaic language, 10 out of 12 interviewees provided information about the general process of formulaic language learning. Six of the 10 were interviewees from the Year 1 group, the rest were from the Year 3 group.

Interviewees from both groups shared a sense that the general process of formulaic language learning, was a linear process. It starts with noticing certain types of formulaic language because they are interesting or well-written, and then these are memorised, followed by a review of formulaic language, and finally ending with the retrieval or recall of the formulaic language at the time of use. This general process is illustrated in example 27. Also, interviewees from both groups mentioned the independent learning of formulaic language, which could assist the formulaic language learning in the classroom. Example 28 provides the opinions of the interviewees on this issue.

Example 27

1. “Firstly, your saying is interesting; you have an interest in it, so you started to learn about it, and then when you master it you can use it in your life (INV 001)”.
2. “*You have to memorise that, for example, this is a good article, then you memorise it, for you may use the language of it later on* (INV 008)”.

Example 28

1. “So, you are kind of independent learning, and then you can learn something new, then put that into the practice, and then study the specific item deeper (INV 001)”.
2. “*I think it is self-learning after the class. You learn it by yourself, and you read, then you can remember that. When it is mentioned by the teachers in the class, you will have a deeper impression of that. It’s like you acquire the knowledge of formulaic language and then practise that in the classroom* (INV 004)”.

However, the interviewees from the Year 3 group added more details to the process of learning, and specially claimed that comprehension has to be achieved before memorising the formulaic language. They also noted the importance of retainment and accumulation in learning formulaic language (example 29).

Example 29

> "*You have to understand the meaning of the language throughout, otherwise it is easy to make a mistake, and for example, it could be the surface meaning rather than the real meaning* (INV 010)".

6.2.2.3 Formulaic language learning strategies

For analysing this code, instead of applying a priori approach to the strategy classification, I use post hoc thematic analysis, which means the interviewees were encouraged to talk about the strategies used in formulaic language learning based on their language learning experience rather than classifying language learner strategies (for more about language learner strategies, see Chapter 4); and then a description of the strategies they preferred to use in formulaic language learning can be provided.

Not all interviewees provided detailed information about the specific strategies that they used to learn the six kinds of formulaic language. Actually, there were only four interviewees (INV 001, INV 002, INV 003, and INV 005) from Year 1 and one interviewee (INV 012) from Year 3 talked about specific strategy use. Therefore, it is impossible to make a comparison or write a description in terms of quantity, or frequency. However, the statements give some information about some of the strategies that they use. Next, I am going to present the analysis of strategies used for each type of formulaic language learning in turn.

For the writing templates, two interviewees (INV 001 and INV 012) mentioned the strategies that they used in learning the writing templates (example 30). They reported that they learned the writing templates from the writing of others, in order to imitate

and use them later; also one interviewee thought that writing templates were quite imperative and needed to be used in writing.

Example 30

1. "The model, I think it is quite imperative. The teacher asks you, you can only write the passage in three parts or five parts. The first part is about your common sense… and other parts are our description and the last part is a summary. The teacher told you that you have to do so. There is nothing have to speak (INV 001)".
2. "*The fixed writing templates is very important. Since only after we know very well about others writing, then we can imitate and use that* (INV 012)".

The second type is idioms and fixed expressions. Each group had one interviewee (INV 002 and INV 012) who reported their learning strategies. For example, they focused on and memorised such types of formulaic language by repeat learning; also, they stated that they needed to expand their repertoire of such types of formulaic language, as illustrated in example 31.

Example 31

1. "*I always paid attention to the first three categories of formulaic language (writing templates, idioms and fixed expressions, and phrases); personally, I likes speech scripts, therefore, he especially focused and learned those kinds of formulaic language from speech scripts, and then read those items out repeatedly; I usually memorised them for later use in writing* (INV 002)".
2. "*Idioms and fixed expressions are important in our daily communication; the more items you have a good command of, the more helpful in the communication. Also, it took some time to remember, once you are familiar with those expressions, you can use them in convenience, it would better if you can put them in sentences or articles* (INV 012)".

Overall, INV 012 seems to have done a lot more work on idioms and fixed expressions than INV 002. Although both interviewees pointed out that they paid attention to the

learning items, and retain some information for later use; INV 012 specifically stated some goals for each step, for example, she noted the importance of enlarging her repertoire of idioms and fixed expressions, increasing her familiarity with the repertoire for using learned items conveniently in communication, as well as the contextualization of the idioms and fixed expressions. It tentatively shows that the interviewee from the Year 3 group understood learning behaviours better than the interviewee from Year 1 group: she knew the reasons for certain learning actions to achieve certain learning goals; also, she knew that language is learned and used in context, not on its own.

As for phrases learning, two interviewees specifically responded with details. Students from both groups responded that only in context, for example, in a sentence, can phrases make sense, as shown in example 32.

Example 32

> "The phrase is only put it into sentences, so it makes sense. If you do not, so it makes nonsense. So you cannot understand it properly. You have to learn some by heart, then you should learn something that you used most frequently by heart. Others you can read by the article or you do some exercise in sentences (INV 001)".

In general, the interviewee (INV 012) from the Year 3 group emphasised the details of phrase learning and noted that they had tried to find better alternative ways to learn efficiently, while the interviewee from Year 1 (INV 001) focused on the linear process of learning, including stages like attention, identification, retention, using learning materials for practicing and contextualization.

As mentioned earlier, phatic expression mainly refers to those expressions that are used fixed and common among people or on one person in daily speech, including greetings, catch-phrases or some small talk in daily speech. For example, the expression *how are you doing?* could serve a phatic function, and performs a social task rather than conveying information in the conversation. Both interviewees knew the function of

phatic expressions, and they acknowledged the role of social interactive strategies in phatic expressions learning (example 33).

Example 33

1. "*When I learn phatic expressions, I will read them aloud few times. After remembering them, I will use them in the communication to get a deeper impression of those phatic expressions…Since phatic expressions are interesting and funny, you can learn them in a recreational and funny ways* (INV 003)".
2. "*I learned those expressions from movie and books, which is context-dependent, and in this way, I said one can learn through your eyes, also know how to use these expressions* (INV 012)".

There was only one interviewee (INV 001) who specifically mentioned his way of learning compound words. In his opinion, the way to learn compound words is actually a way of understanding word formation, and he expressed that in example 34.

Example 34

> "You can learn a word by word formation. So, you can say that one word, and through the word formation method, you can learn three or four more words. You can change a word, or you can add some elements, as it is changed into another word. So it is very efficient to learn words by words formation (INV 001)".

As for the terminologies, two interviewees (INV 001 and INV 012) reported the strategies that they used in learning this type of formulaic language (example 35). They thought that terminology was related to a specific field, and involved detailed knowledge of that field. Therefore, when learning terminology, one needs to put it in the context of the specific field.

Example 35

1. "I am learning vehicle in English, you have to give a text about the vehicle. So I can learn some specific words in that article. Put the phrases and words into

sentences of the article, so that it makes sense. Maybe, I don't know it is right or not. It is just my way (INV 001)".

2. "*I and my classmates were poor-motivated to learn terminologies or compound words, only depended on the teachers' instruction, it just makes things easier* (INV 012)".

Only five out of 12 interviewees provided specific information on the strategies that they used to learn each kind of formulaic language. However, it is still rational to draw a conclusion from these interviewees' statements analysed above as follows:

First, not all interviewees were able to report specific strategies for each kind of formulaic language learning, which could indicate that they were not aware of some strategies used in learning. Also, this may be caused by a lack of explicit teaching or instruction from their teachers. Second, it shows that, in general, the interviewee from Year 1 tended to focus on the linear process of language learning, including attention, identification, retention, and saving learned language for later use; whereas the interviewee from Year 3 seemed to have a more comprehensive view of language learning, and made comparisons between different strategies, for example, learning in context and without context. Third, interviewees from both groups realised that these six kinds of formulaic language are context-dependent and it is better and more efficient to put language in context. Last, guidance from teachers is needed when affective factors arise in learning.

In this section, the data regarding formulaic language learning from two rounds of interviews were examined and analysed. With regard to the general learning aspect, interviewees from both groups shared the view that the general process of formulaic language learning was linear. It started with noticing, then memorisation, followed by the review or retainment of formulaic language and ended with the retrieval or recall of the formulaic language at the time of use. However, the interviewees from the Year 3 group reported a broader range of strategies than the Year 1 group for formulaic language learning. As for the learning sources, both groups relied on visual learning materials. The interviewees from the Year 1 group referred to dictionaries or foreigners

as language learning resources to confirm the correctness of their language, whereas the Year 3 group interviewees perceived themselves to be the language users. They tried to use English and referred to the different available resources rather than only focusing on correctness in their language production in terms of formulaic language. By asking if formulaic language is important in language learning, the answers from the interviewees show that not all interviewees thought that formulaic language was important for English language learning, and all interviewees had an awareness of formulaic language during their learning of English.

For learning specific types of formulaic language, in terms of learning preferences, among the six categories, phrases, and idiom and fixed expressions rank at the top of the list for interviewees from both groups. The interviewees from Year 1 give priority to learning writing templates and compound words, whereas those from Year 3 would like to learn some formulaic language like phatic expressions and terminology. As to the strategies, not all interviewees were able to report specific strategies for each kind of formulaic language learning, which could indicate that they were not aware of some strategies used in learning. It also shows that the interviewees from Year 1 tended to focus on the linear process of language learning, including attention, identification, retention, and saving learned language for later use; whereas for the interviewee from Year 3, seems to have a more comprehensive view of language learning, as she made a comparison between different strategies. In addition, interviewees from both groups realise that these six kinds of formulaic language are context-dependent and it is better and more efficient to put language in context. Last, guidance from teachers is needed when affective factors arise in learning.

6.2.2.4 Formulaic language learning sources

As for the sources the interviewees used for learning formulaic language, the following two tables show the types of resources which were mentioned by the interviewees from both groups in the interview. Table 6.25 shows the responses from the interviewees in

Year 1, and Table 6.26 shows the information provided by the interviewees from the Year 3 group.

Table 6.25: Information on the resources used by the Year 1 interviewees in formulaic language learning

Resources	**Interviewees**
real/ authentic/original English article	INV 001
dictionary	INV 001, INV 002, INV 006
teachers' instructions	INV 001, INV 006
self-monitor	INV 001
communication with foreigners (for confirmation)	INV 001
books/articles written by English people	INV 001
textbooks	INV 002, INV 004, INV 005
movies in English	INV 002, INV 005, INV 006
books/magazines in English	INV 002, INV 006
all sorts of reading in English outside of the classroom	INV 002
support materials for exams	INV 004
online blogs/post in English	INV 004
TV Drama in English	INV 005
Handouts given by the teachers	INV 006

Table 6.26: Information on the resources used by the Year 3 interviewees in formulaic language learning

Resources	**Interviewees**
all sorts of literary works	INV 001
newspapers	INV 007
poems	INV 007
textbooks	INV 007
classroom learning	INV 008, INV 009
any English texts	INV 009
English radio (BBC)	INV 009
English books/articles/magazine	INV 009, INV 010
English movie/TV drama/	INV 010, INV 012
fairy tales in English	INV 012
foreigner teachers (for communication)	INV 012
peers' communication	INV 012
fictions	INV 012

From the table above, it is clear that the interviewees from the two groups had the same or similar resources for formulaic language learning, such as movies, TV dramas,

textbooks, reading. Almost all of these resources were introduced by visual learning methods, in other words, the interviewees need to watch or read.

The interviewees from the Year 1 group referred to dictionaries or foreigners as their language learning resources to confirm the correctness in their language, whereas in the Year 3 group, the interviewees perceived themselves as the language users, for instance, they communicated with foreigners in English, in other words, they tried to use English and referred to different available resources rather than only focusing on the correctness in their language production in terms of formulaic language. Also, it seemed that the interviewees from the Year 1 tended to learn anything in English; however, the interviewees from the Year 3 group tended to aim for something from original, authentic English, rather than just considering any material in English as a resource for learning. Apart from this, it is noted that the interviews did not specify the differences of formulaic language in the written or spoken register.

6.2.2.5 Learning preferences

In the second round of interviews, the list of six kinds of formulaic language above was given to the interviewees[12]. All the interviewees were then asked to rank these six categories according to their preference within a range from 1 to 6 (1 means the most preferred item, whereas 6 means the least preferred item) in their own English language learning.

For analysing the data regarding the learning preferences in formulaic language, they were put as variants in SPSS (Table 6.27).

Table 6.27: Overview of learning preference in the six types of formulaic language

interviewees	Phrases	Compound words	Terminology	Idioms and fixed expressions	Writing templates	Phatic expressions

[12] The interviewee mentioned collocations in the interviews. However, collocations were ruled out in this research for the two reasons: first, there is no consensus on whether collocations can be regarded as formulaic language; second, collocations reflect co-occurrence nature, which is not in the research scope.

INV 001	1	2	5	3	4	6
INV 002	3	5	6	2	1	4
INV 003 (1)	2	5	4	1	3	6
INV 003 (2)	4	3	5	2	6	1
INV 004	2	3	4	5	1	6
INV 005	3	4	5	2	6	1
INV 006 (1)	1	4	5	2	3	6
INV 006 (2)	3	4	6	2	5	1
INV 007	4	5	1	3	6	2
INV 008	4	6	3	2	5	1
INV 009	1	3	5	2	4	6
INV 010	1	3	5	2	4	6
INV 011	1	3	4	5	3	6
INV 012	4	5	6	2	2	1

In the table above, there are two interviewees (INV 003 and INV 006) who gave alternative preferences depending on different learning purposes and situations.

Numbers 1 to 6 were assigned to each type of formulaic language and stand for the order given to each category by the interviewees. For example, interviewee 007 gave top priority to learning terminology, followed by phatic expressions, idioms/fixed expressions, phrases, compound words and writing templates.

Next, based on the frequency in the descriptive analysis in SPSS, an overview of learning preferences among the six types of formulaic language is presented (Table 6.28).

Table 6.28: Learning preferences of formulaic language from all interviewees

Total number	**Phrases**	**Compound words**	**Terminology**	**Idioms and fixed expressions**	**Writing templates**	**Phatic expressions**
Interviewees	14	14	14	14	14	14
Preferences	34	55	64	35	53	53

In the table above, all 12 interviewees ranked six types of formulaic language, and two interviewees (INV 003 and INV 006) had two sets of rankings. In order of preference, it shows that the higher the score that the item gained, the lower the preference in terms of formulaic language learning. Therefore, it is clear that “phrases” hold the first place, followed by “idioms and fixed expressions”, and then “writing templates” and “phatic

expressions" share the same priority in the preference list. "Compound words" come afterwards, and "terminology" is last.

The tables below show the learning preferences among these six types of formulaic language ranked by the interviewees from both groups respectively (Tables 6.29 and 6.30). Again, the analysis results from the frequency analysis in SPSS.

Table 6.29: Year 1 group learning preferences

	Phrases	**Compound words**	**Terminology**	**Idioms and fixed expressions**	**Writing templates**	**Phatic expressions**
INV 001	1	2	5	3	4	6
INV 002	3	5	6	2	1	4
INV 003 (1)	2	5	4	1	3	6
INV 003 (2)	4	3	5	2	6	1
INV 004	2	3	4	5	1	6
INV 005	3	4	5	2	6	1
INV 006 (1)	1	4	5	2	3	6
INV 006 (2)	3	4	6	2	5	1
Sum: 8	19	30	40	19	29	31

Table 6.30: Year 3 group learning preferences

	Phrases	**Compound words**	**Terminology**	**Idioms and fixed expressions**	**Writing templates**	**Phatic expressions**
INV 007	4	5	1	3	6	2
INV 008	4	6	3	2	5	1
INV 009	1	3	5	2	4	6
INV 010	1	3	5	2	4	6
INV 011	1	3	4	5	3	6
INV 012	4	5	6	2	2	1
Sum: 6	15	25	24	16	24	22

In Table 6.29, two interviewees (INV 003 and INV 006) ranked the items twice, thus the sum of the ranking times is eight. As for the other six columns, the smaller the number is, the more it is preferred by the interviewees. So, it is shown that for the interviewees from Year 1, their learning preferences are as follows: in first place, phrases, and idioms and fixed expressions, writing templates, compound words, phatic expressions, and

terminology; whereas the Year 3 group interviewees give priority to phrases, idioms and fixed expressions, and phatic expressions, as the first three. Then the terminology and writing templates share the same placement in the preference table and compound words come as the last item.

Among these six categories, phrases, idioms, and fixed expressions rank at the top of the list for interviewees from both groups. The interviewees from Year 1 give priority to learning writing templates and compound words, whereas those from Year 3 would like to learn some formulaic language like phatic expressions and terminology.

Also taking the whole picture into consideration, it is easy to spot that not all kinds of formulaic language were ranked across all placements by the interviewees. For example, the category of phrases, and idioms and fixed expressions was never put in the last place of the list, while the category of compounds was not put as the top priority ever. On the basis of this, the assumption could be that for the interviewees, the category phrases, and idioms and fixed expressions are more important than the category of compound words in terms of formulaic language learning.

6.2.3 Using formulaic language

Under this code, two main themes emerged and were grouped in the sub-categories as the process and strategies in using formulaic language. In this section, the focus is put on formulaic language in use, rather than the learning process. Next, I will present and analyse the information concerning these two codes.

6.2.3.1 Process

In total, 10 out of the 12 interviewees from both groups talked about the process of using formulaic language. There were six interviewees from the Year 1 group, and four interviewees from the Year 3 group.

According to the statements in the interviews, interviewees from both groups shared a similar view of the process of using formulaic language and reported that it involved the transfer from English to Chinese, and then Chinese to English (Example 36).

Example 36

1. “In Chinese, if describe a thing, we need certain methods, but in English, we don't. For example, don't be on my way, if we think in Chinese, it means don't stand on the road, and then transfer that to block someone’s way. But we don't need that kind of transfer in English (INV 001)”.
2. “*Based on the previous Chinese expressions, I will find their equivalent phrases or the synonyms, and then use them* (INV 011)”.

According to the statement above, the process of using formulaic language is depicted in Figure 6.22.

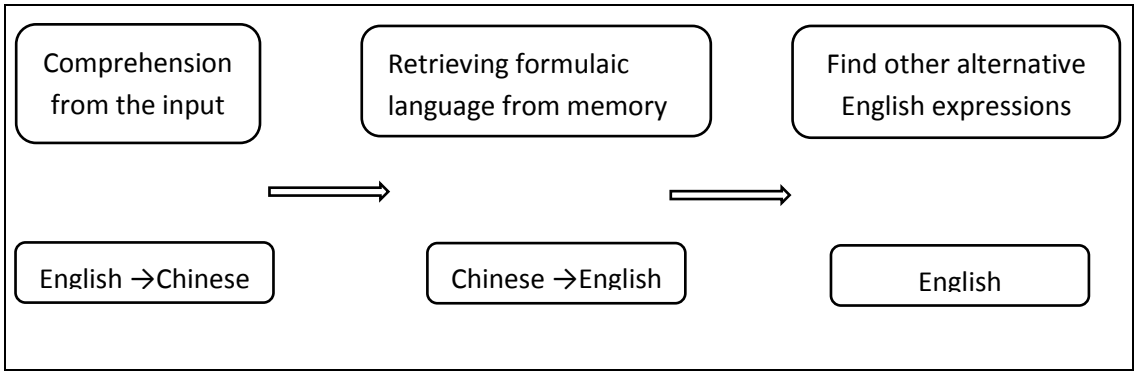

Figure 6.20: The process of using formulaic language

From the figure above, the upper labels are the stages in the process of using formulaic language. According to the interviewees, it started with the comprehension of input materials, at this stage, the interviewees needed to transfer the language of the input materials (English) to their L1 (Chinese) to comprehend them. This was followed by the process of retrieving formulaic language in the memory, usually the interviewees find equivalent English expressions which match the comprehended Chinese expressions. Sometimes, the interviewees would find more alternative English expressions.

6.2.3.2 Strategies

Strategies for using formulaic language are intertwined with the process of formulaic language use, and there was not a clear-cut distinction between the process and strategies for using formulaic language. The process often consists of continuous evolving stages, whereas the strategies can be more specific and used to solve certain issues. In this part, I will present and examine the data extracted from the interviews to detect the strategies that applied to formulaic language.

All 12 interviewees from both groups gave some information about the strategies claimed that they applied when they used formulaic language. In general, they talked about the use of strategies with three aspects. Next, I will analyse them in turn.

First, in the formulaic language using process, interviewees reported that language transfer is used as an analytic tool. On one hand, once they have the Chinese expression, they then seek for the equivalent in English from memory, as illustrated in example 36. On the other hand, by the comparison between two languages, the students found language items that were different from their L1 (example 37).

Example 37

> "*For example, he gave two examples, for example, for instance, and then normally, you would not give too many examples, four at most. You can use the reason are as follows, or you can use the word reason, you can replace reason with other words, or, you can use such as, or just use colons, or e.g. or something else* (INV 004)".

Example 38

> "*Hug you into pieces, for example, this phrase is interesting, and in Chinese, there is no such expression. I may remember it easily* (INV 011)".

Second, interviewees claimed that they used contrastive rhetoric as a strategy in using formulaic language. The interviewees talked about comparisons with Chinese writing and its influence on their English writing, for example, they claimed that they tried to

avoid common formulaic language use in order to be creative in English writing. This view is expressed in example 39.

Example 39

> "*In the high school, the teachers told us it was not good if you used some formulaic language or words again and again in your composition, it would make your audience sleepy; Instead, you have to use formulaic language in different ways and make some changes to that* (INV 005)".

However, the general solution to the issue of using the same phrases or collocations in a text repeatedly, was proposed by three interviewees (INV 001, INV 002, INV 003, and INV 004). This was to find more alternative expressions to replace those repeatedly used phrases or collocations (example 40).

Example 40

1. "*I just look for the synonyms or change the form of the words* (INV 002)".
2. "*I just make some changes. I use these expressions earlier and use the other one later on. Usually, I use some formulaic language with same meaning, but the different structure or words in the expression* (INV 008)".

As the third point gained from the interviews, students from both groups also talked about what they did and did not do when using formulaic language. In terms of "what they did" when using formulaic language, interviewees reported that they used formulaic language and writing templates in different ways depending on the different purposes of the writing, for example, they only used writing templates in English writing for assessments, not for personal purposes (example 41).

Example 41

> *"I think, if write for exams, I will use some models of writing, which is already formed in your brain. In this way, you can get higher scores. While, I don't think it I have to obey those rules in the exams when I write for dairy. Different from in the composition, in dairy writing, you do need to think about the fixed collocations, or fixed sentence types. You can write relatively freely. But in*

composition writing, you cannot do this. You have to think about using some fixed phrase, sentence types and fixed collocations. They are two different kinds of feeling (INV 003)".

As for the "what they did not do", two interviewees (INV 004 and INV 005) mentioned famous quotes, idioms and fixed expressions, and large chunks of language from the writing of others (example 42).

Example 42

1. *"Sometimes he changed parts of the formulaic language, but he never changed any word in the famous quotes* (INV 004)".
2. *"I barely used the memorised long sentences in the English writing', instead, he expressed his ideas in his own words* (INV 005)".

In addition, students also stated that using formulaic language aims to achieve "the varied, elegant, native and creative use of English language (INV 008 要变化，要典雅，要地道，要有新奇感)" and avoid the repeated use of formulaic language, especially in writing. They gave the following reasons for this. For example, first, the students thought that the repeated use of formulaic language in one text made the text boring and tedious; second, they worried that the repeated use of formulaic language reflected a limited English proficiency level in the writer (see example 50). However, they seemed not to be bothered by formulaic language used in writing to their peers', as they just tried to find their own personal way of using formulaic language (example 43).

Example 43

1. "*If you use some formulaic language repeatedly in a text, it looks tedious, and the audience may feel the writer lacks creative thinking* (INV 012)".
2. "*If you use some phrases all the time, people may think your English level is limited. If you really are good at English, you can use formulaic language in various ways, you can choose other ways of expressions. If you use the same formulaic language in the same form, it only showed your limited vocabulary* (INV 012)".

Example 44

> "*I was worried sometimes, but you do not need over-worried about that. For some formulaic language, you had no way to stop others using it. However, you have to use that in your own ways, for example, you can use different places in your writing, or use in different occasions* (INV 012)".

In a nutshell, this section presented the data and analysis of the process and strategies for using formulaic language. As a continuous process, there seemed to be no clear-cut distinction between using formulaic language and learning formulaic language, or between the process and strategies involved in using formulaic language. In fact, they all tended to be intertwined with each other.

Interviewees from both groups shared the general process of using formulaic language, which went through comprehension of the input materials, retrieval of formulaic language from memory and sometimes, finding more alternatives to realise the varied or creative use of formulaic language. For the strategies that interviewees claimed to use when using formulaic language, it is noted that the interviewees from both groups realised that there was a process of transfer between two languages in using formulaic language. Most times, it is used it as an analytic strategy to help learners find gateway to remembering formulaic language, for example, when it is different from their L1, when it is easy to recall or retrieve from the memory, or to help to establish links between Chinese expressions and the English equivalents.

6.2.4 Students' comments on formulaic language teaching

In this section, the comments from the interviewees on formulaic language teaching from the two rounds of interviews are presented. The interviewees from both groups were invited to talk about the teaching of formulaic language that they had received since they studied in middle schools. This theme emerged from the second round of interviews. Under this theme, the interviewees were invited to talk about the teaching of formulaic language that they had received. The content can be categorised into three subcategories: formulaic language teaching, textbooks and teachers. Next, an analysis of these themes will be presented in turn.

6.2.4.1 Formulaic language teaching on research site

Under this theme, the focus is on finding out what teachers teach in the classroom regarding formulaic language and how formulaic language is taught at the research site. Thus, all the 12 interviewees were asked questions like “what do your teachers teach you about formulaic language in the classroom?” “Do they introduce any strategies regarding formulaic language learning and use?”.

With regard to the issue of how formulaic language is taught, students reported that most of the teaching regarding formulaic language employed a grammar-translation approach with teachers giving explicit instructions on the analysis of grammatical features of formulaic language (example 45).

Example 45

> “*Take any article in the textbook as an example, teachers always gave us a simple analysis of the sentences or phrases, and then we can comprehend that by ourselves, mainly outside of the classroom. The analysed parts include some difficult sentences involved lots grammar stuff or certain sentences with the special structure or some paragraphs* (INV 006)”.

The analysis shows a more complex picture when it comes to the issue of what teachers teach regarding formulaic language. First, it was reported that teachers at the research site introduced different types of formulaic language to the students in the classroom (example 46). Second, the teachers also introduced some learning strategies (although most of them are mnemonics techniques, such as classification and association or they devised some assessments based on the memorisation of formulaic language, for example, dictation (Example 47).

Example 46

> “*My teacher introduced that they had to learn the writing templates, compound words and terminologies by rote learning, in order to memorise these types of formulaic language accurately and mostly, writing templates would be used in*

the writing exams, for example, TEM 4; as for the phatic expressions, we could learn that in the pre-scripted dialogues (INV 007)".

Example 47

"*Teachers introduced some techniques for memorising formulaic language, such as using association or classification of the items. Also, we were kind of forced to memorise the formulaic language that the teachers taught in the class since we have to do the dictation on the formulaic language as a means of assessment. You have to do that* (INV 003)".

In the interviews, the students were also asked to think in retrospect about formulaic language teaching across different language learning levels, from middle schools to high school to university. The evidence from the analysis shows that the teaching of formulaic language at each stage had different focuses, and the teaching at the university was a watershed: there was more instruction on formulaic language in the English language classroom in middle schools and high schools (Example 48).

Example 48

1. "*Formulaic language teaching in the university is not like that in the school, for instance, like before the university, in the basic language learning stage, the teachers emphasised that we should memorise some kinds of formulaic language, such as fixed expressions or idioms and certain sentence types. However, now, the focus is put on the use formulaic language in a flexible way* (INV 009)".
2. "*When passing exams become the purpose of learning English, teachers would like to pay more attention to the language usage that appears in the exam papers. For example, the conjunctional words in the passage, or some "good beginning sentences" for a composition, and some proverbs, they will think of the benefits that these kinds of language will bring to the writing. Also, readings in the exams gain a lot of attention, for example, how students could comprehend the passages by learning these kinds of language. But once, you entered into the university, nobody asks you to learn this kind of language, and it totally becomes a kind of interest, out of some academic drive* (INV 001)".

First, two interviewees (INV 002 and INV 003) claimed that their teachers emphasised the importance of formulaic language, but did not give any explanation of functions that made formulaic language important in language learning. This view is illustrated in example 48. Second, interviewee (INV 006) also claimed that in terms of formulaic language teaching, her teachers demonstrated the strategies used in learning formulaic language, as she expressed it in example 54. Thirdly, one interviewee (INV 001) stated that there were some differences between the formulaic language that he was taught by teachers in class and the language that he learned by himself outside the classroom (example 49). Furthermore, this statement again reflects the complicated relationship between spoken and written formulaic language learning in the research context.

Example 49

> "*Since it is very hard to learn in books or learn it from teachers, some usage that if you want to speak like the English, so you have to read them in articles, read articles written by them. For example, if you want to learn some beautiful words, you have to read something like Shirley, or something from Shakespeare, if you cannot, so you rarely learn idiomatic usage in English, maybe it is Chinglish* (INV 001)".

In the Year 3 group, three (INV 007, INV 008 and INV 010) out of six interviewees talked about their opinions on formulaic language teaching. They noted the different types of formulaic language that they could be learned in the classroom and some that they could probably only learn outside the classroom, as an interviewee (INV 007) expressed in example 51. Also, interviewee (INV 007) mentioned that their teacher had introduced some techniques for memorising formulaic language, such as using pre-scripted dialogues or memorising the context (example 52). In addition, interviewees (INV 009 and INV 010) said that their teachers taught formulaic language differently in terms of content according to the different stages that they were at (example 50).

Example 50

"*We can learn some kinds of formulaic language, like terminologies, idioms and fixed expressions in the classroom, whereas some other types of formulaic language are important in speaking, but we have to learn those by themselves outside of the classroom* (INV 008)".

Example 51

"*As writing template, when we prepared for the TEM-4 or earlier in the high school, teachers taught us the three-paragraph essay, we can use that for writing in the exams; also, we were told certain kind of language we have to remember, for instance, phrases. Something like greetings or phatic expressions, teachers asked us to memorise the dialogue in the textbooks. The rest, compound words, and terminologies, we learn through rote learning* (INV 007)".

Example 52

"*There are substantial differences between the formulaic language teaching in the university and at schools, since the university teaching tends to focus on the understanding of the whole text or paragraph, or the underlying meaning that the writer wanted to express; rather than concentrating on specific language use details* (INV 010)".

To conclude, in general, the students reported that they were satisfied with their teachers' teaching content regarding formulaic language; however, the interviewees also had remained their own opinions about formulaic language teaching. For example, interviewees from both groups mentioned that certain types of formulaic language that they learned in the class could be used in certain contexts, but that they needed to learn some "real or authentic English" in order to use it in speaking, and they could not learn this type of "real or authentic English" in the classroom. Also it can be seen that there are some changes in teaching content for Year 1 and Year 3 students: more instructions on grammar and detailed language analysis was put in Year 1 teaching, while the teaching for the Year 3 students, paid more attention to general language use, and meaning-making. With regard to the teaching content, the interviewees from Year 1 said that their teachers stressed the importance of formulaic language in the class, however,

teachers did not make it clear whether formulaic language was important for language learning or for students to pass exams. This vague instruction left students without a good explanation of the function of formulaic language.

6.2.4.2 Textbooks

As an important part of the teaching, the textbook for each module is fixed and chosen by the lecturers and the department at the research site. Each module, in general, has one textbook. *Family Album U.S.A.*, for example, is the textbook that they use in the compulsory module *Advanced Viewing, Listening and Speaking* for all the Year 3 students in the department. In other words, the content of the textbook decides the content that is taught and learned in the classroom (see Chapter 5 for more details).

11 of the 12 interviewees gave their comments on the textbook content regarding formulaic language. Five of the six interviewees from Year 1 gave their opinions on the role of the textbooks in formulaic language teaching and learning; whereas all the interviewees from Year 3 commented on the issue.

For these five interviewees from Year 1, four of them made positive comments about the content of textbooks meeting their formulaic language learning needs. They used words such as 'helpful' (INV 003), 'satisfied' (INV 002 and INV 004), and 'excellent' (INV 005). They also gave the reasons for these judgements (example 53).

Example 53

> "*They could influence the students in an implicit way and make them think in English way, as well as cover the language use in exams and daily use* (INV 003)".

There was one interviewee (INV 006) who was not satisfied with the textbooks since she thought the content of the textbooks was quite limited, as expressed in example 54.

Example 54

"*The content in the textbook is quite limited, and thus probably one cause of the repetition or repeated use of some formulaic language, especially in the English written texts in the class* (INV 006)".

With respect to the Year 3 students, they seemed to give more varied answers to this question. Only one interviewee (INV 009) claimed that the content of the textbooks satisfied his learning needs (example 65). One interviewee (INV 012) said that she had never paid attention to the role of the textbook in formulaic language teaching and learning. Four interviewees gave negative feedback on the content of textbooks for meeting their formulaic language learning needs. Two of them (INV 008 and INV 010) said that the content about formulaic language in textbooks was "not much or enough"; the other two interviewees (INV 007 and INV 011) made it clear that they were not satisfied with the content of formulaic language in their textbooks, as the content was too limited (example 55).

Example 55

"*I think they can meet my learning needs, as they cover most areas that I would use English, especially in terms of formulaic language, especially beneficial for the future professional English language use* (INV 009)".

Example 56

"*The content of textbook was limited, and students cannot learn all six types of formulaic language only from the textbooks; for example, the textbook for Advanced English, the interviewee reported that most of the articles they learned in that book are literature works; you cannot find any writing templates we used in the exams from the textbooks; and some formulaic language can be used in current colloquial context or in real life* (INV 007)".

Comparing the comments from both groups, it can be concluded that the Year 3 interviewees considered textbooks to be less influential since they presented more varied opinions and attitudes on the textbooks they used for learning formulaic

language. Instead of being simply satisfied or not, they also gave the reasons for their comments.

6.2.4.3 Teachers: foreign and domestic teachers

In the second round of interviews, the 12 interviewees were also asked to comment on formulaic language teaching in terms of teachers. According to the statements from interviewees, I divided these comments into two aspects: students' statements on domestic and foreign teachers.

11 interviewees commented on their domestic teachers in terms of their formulaic language teaching; however, only three Interviewees (INV001, INV 005 and INV012) talked about their foreign teachers. On one hand, the reason for this could be that the interviewees received far more instruction in the English language classroom from domestic teachers than from foreign teachers, especially for the Year 3 students, as they did not have English speaking classes led by foreign teachers; on the other hand, most students did not have foreign teachers until they entered university, even in the university, foreign English teachers played a complementary part in English teaching at research site. Next, I will examine and analyse the statements from the interviewees on formulaic language teaching with regard to domestic and foreign teachers.

As for the domestic teachers, the students' comment on teaching mainly focused on the issue of what can be taught regarding formulaic language. On one hand, a student raised questions about what type of formulaic language should be taught in the classroom, as the student thought that formulaic language was over-taught in the university, especially for the "verb + n" phrases or collocations, as "it is common", and it is transparent in the meaning of the phrase; whereas he expressed that he needed more instruction on writing templates and good sentences for opening and closing an English composition (example 57).

Example 57

1. "*University teachers, I think they sometimes over-emphasize the formulaic language learning in their teaching. For example, the V+ N phrases. I don't think they are so called fixed phrase, or formulaic language. And then, teachers asked us to learn the phrase play table tennis, by heart, as a fixed phrase. I personally feel if it is common in the books, then, it is not a fixed phrase* (INV 004)".
2. "*The teachers should give us some more exercise for practice, including some good opening sentences and ending a sentence in the English writing, some good writing samples with comments alongside, some writing templates. I think it is better for improving our ability and levels in English writing* (INV 004)".

It is noted that some formulaic language teaching was not accepted by the students, or in other words, common ground regarding formulaic language teaching had not yet reached between domestic teachers and the students.

On the other hand, students from both groups realised that certain types of formulaic language that they learned in class could be used in certain contexts, but that they needed to learn some "real or authentic English" in order to use it in speaking, and they could not learn this type "real or authentic English" in the classroom (example 58).

Example 58

1. "Since it is very hard to learn in books or learn it from teachers, some usage that if you want to speak like the English, so you have to read them in articles, read articles written by them. For example, if you want to learn some beautiful words, you have to read something like Shirley, or something from Shakespeare, if you cannot, so you rarely learn idiomatic usage in English, maybe it is Chinglish (INV 001)".
2. "*We can learn some kinds of formulaic language, like terminologies, idioms and fixed expressions in the classroom, whereas some other types of formulaic language are important in speaking, but we have to learn those by themselves outside of the classroom* (INV 008)".

In addition to this, one interviewee (INV 006) reported that she was a little bit worried about the method of teaching formulaic language in larger size classes (example 59).

Example 59

> "*I was thinking, if everyone in the class received the same hand-out of formulaic language from one teacher, I was afraid of that everyone would use the same language, especially in the writing. I then started to think about changing the expression, or create a new way to express myself* (INV 006)".

It is noted that generally instruction in formulaic language is given by domestic teachers across all levels, from middle school to high school to university. However, the extent that domestic teachers included formulaic language in their instruction decreased gradually, from middle school, which is where English learning begins, to university. Interviewees also retained their own opinions on formulaic language teaching. A consensus on the question "what types of formulaic language should be taught in the classroom", was not reached by students and teachers. Moreover, they never talked about this issue to each other. It is also noted that domestic teachers were influenced by native norms, and as the interviewee reported, the native norms were simplified by using varied expressions.

The foreign teachers, since they did not play a leading role at the research site, had less time for language teaching. Still, three interviewees gave their comments on their teaching regarding formulaic language.

One of the interviewees (INV 001) noted that his foreign teachers encouraged him to learn and use idiomatic expressions in his English and that idiomatic usage was promoted as one of the merits of good English writing (example 60). The second Interviewee (INV 005) gave an example of how foreign teachers criticised the content of textbooks on formulaic language (example 61). The third interviewee (INV 012) reported that implicit teaching probably happened in the communication with her foreign teachers on campus (example 62).

Example 60

> "And my teacher Daniel often said that "you are an enthusiasm writer, however your writing is lack of idiomatical convey and what you should do is read more articles written by the native English speakers, and it's also very important to learn some words by heart and some idiomatical usage learned by heart (INV 001)".

Example 61

> "*Once criticised the English Speaking textbook, as the word different used repeatedly across a dialogue and should be changed into other words, for example, differ from; by means of this, the language sounds not that boring* (INV 005)".

Example 62

> "*Through the interaction or communication with the foreign teachers, since they are the native speakers of English, and I can learn the real and authentic English from them. Rather than you learn the language from the textbooks, you can really communicate with the foreigners, both in and outside of the classroom. I learned a lot* (INV 012)".

To sum up, in terms of domestic and foreign teachers and formulaic language teaching, it is noted that first, the foreign teachers were strongly influenced by native norms in their teaching and encouraged students to use idiomatic expressions explicitly, even in their personal communication. Second, both domestic and the foreign teachers told students to avoid using the same expressions or phrases all the time in one text, and instead to use the different word formations or synonyms. However, the foreign teachers were subtly different from the domestic teachers in their attitudes towards the textbooks. The foreign teachers, as reported, criticised the content of the textbooks and tended to be more confident in language use than the domestic teachers. Third, foreign teachers were recognised by the students as representative of native and authentic English speakers. Additionally, it is worth noting that the foreign teachers tended to focus more on "communication", like personal communication and real communication in English. Also, formulaic language teaching was applied in both implicit (conversation

with native speakers) and explicit (instructions in the classrooms and personal advice) ways by foreign teachers.

6.2.5 Analysis of other contextual elements on the research site

In this section, I analysed the issues raised by the interviewees in the second round of interviews in two aspects: internal and external factors.

For the internal factors, I gathered the information provided by the interviewees, presented and analysed it in terms of English language motivation and reflections on English language learning. In the analysis of language learning motivation, it was noted that most students had a mixture of integrative and instrumental motivation, and that they tended to swing between these two. With regard to the reflections on language learning, I divided the information into two groups: reflections on language performance and language learning. The interviewees had high expectations of their English language performance and did not want to show their shortcomings; as for the English learning, they realised that the lack of immersion in context and accumulation in learning, as well as their relatively passive learning attitudes led to various challenges or so-called "bottleneck" in their English language learning.

As for the external factors, I mainly examined the institution's policies and influence (including the construction of facilities and equipment on campus for English language learning, as well as the syllabus design), English language teachers and students' peers as possible factors that influence the students' English learning from the outside. The imbalance in the development policy and strategies for different subject disciplines was reflected in the analysed aspects. Factors like shortages of equipment and facilities for English language learning, the urgency to improve the quality and quantity of teaching staff (which has already influenced syllabus design), some pressure from peers and the low level of willingness to communicate in English, played a negative role in English learning at the research site.

6.2.6 Summary of interview analysis

With regard to the understanding of formulaic language, in general, interviewees from both groups reported fixedness as a more important feature of formulaic language than frequency, and they also stated that writing templates are a type of formulaic language, as they are important in English writing. Also the interviewees from Year 3 talked more broadly about the possible forms of formulaic language, such as greetings, proverbs, and famous quotes. In terms of function, the interviewees regarded formulaic language as a yardstick that L2 learners have to follow since it can guarantee language use within the requirements. Also, it was convenient to use ready-to-use formulaic language when needed, as it could be easily and quickly retrieved from the memory, which saved time and effort by avoiding the construction of new language every time. In addition, they contended that formulaic language could be a great addition to learners' language, since it can increase grammatical correctness. Some interviewees also claimed that sometimes, formulaic language can be used for producing funny or humorous effects. One interviewee also provided a description of the changing role of formulaic language during the language proficiency developmental process.

As for the understanding of specific types of formulaic language, the interviewees showed that they were familiar with writing templates, and talked about them in terms of form and function. However, for the rest of the five types of formulaic language, the interviewees were only able to give rough descriptions.

The two rounds of interviews examined formulaic language learning reported by the interviewees. With regard to the general learning aspects, interviewees from both groups shared the general process of formulaic language learning, which was a linear process. It started with noticing, and then memorisation, followed by the review or retainment of formulaic language, finally ending with the retrieval or recall of the formulaic language at the time of use. However, the interviewees from the Year 3 group reported broader strategies than the Year 1 groups for formulaic language learning. As for the learning sources, both groups relied on visual learning materials, and the

interviewees from the Year 1 group referred to dictionaries or foreigners as language learning resources to confirm the correctness in their language, whereas the Year 3 group interviewees perceived themselves as the language users. They tried to use English and referred to the different available resources rather than only focusing on the correctness of their language production in terms of formulaic language. By asking if formulaic language is important in language learning, the answers from the interviewees showed that not all interviewees thought that formulaic language was important for English language learning, and not all interviewees had an awareness of formulaic language during their English learning.

For learning specific types of formulaic language, in the learning preferences, among the six categories, phrases and idioms and fixed expressions rank at the top of the list for interviewees from both groups. The interviewees from Year 1 give priority to learning writing templates and compound words, rather than those from Year 3: who would like to learn some formulaic language like phatic expressions and terminology. As to the strategies, not all interviewees were able to report specific strategies for each kind of formulaic language learning, which could indicate that they were not aware of some strategies. It also shows that the interviewees from Year 1 tended to focus on the linear process of language learning, including attention, identification, retention, and saving learned language for later use; whereas the interviewee from Year 3 seemed to have a more comprehensive view of language learning, as she made comparisons between different strategies. In addition, interviewees from both groups realised that these six kinds of formulaic language are context dependent and it is better and more efficient to put language in context. Last, guidance from teachers is needed when affective factors arise in learning.

The process and strategies for using formulaic language were also analysed in the two rounds of interviews. As a continuous process, there seemed to be no clear-cut distinction between using formulaic language and learning formulaic language, or between the process and strategies for using formulaic language. In fact, they all tended

to be intertwined with each other. For the process of using formulaic language, the interviewees from both groups already realised that they used formulaic language in English writing, however, there was no strong evidence showing their awareness of using formulaic language in speaking, or in their L1, Chinese. Interviewees from both groups shared the general process of using formulaic language, which went through the stages of comprehending the input materials, retrieval of formulaic language from memory, and sometimes, the use of more alternatives to realise the varied or creative use of formulaic language. For the strategies that interviewees claimed to use for formulaic language, it is noted that the interviewees from both groups realised that the process of transfer between two languages involves formulaic language. Most times, it is used it as an analytic strategy to help learners find the gateway to remembering formulaic language, for example, when it is different from their L1, it was easy to recall or retrieve from the memory, or helped to establish links between Chinese expressions and the English equivalents.

In terms of formulaic language regarding domestic and foreign teachers, it is noted that first, foreign teachers were strongly influenced by native norms in their teaching and encouraged students to use idiomatic expressions explicitly, even in their personal communication. Second, domestic teachers and foreign teachers told students to avoid using the same expressions or phrases all the time in one text, and instead to use the different word formations or synonyms. However, the foreign teachers were subtly different from their domestic teachers in the attitudes towards textbooks. The foreign teachers, as reported, criticised the content of the textbooks and tended to be more confident in language use than the domestic teachers. Third, foreign teachers were recognised by the students as representative of native and authentic English. Additionally, it is worth noting that the foreign teachers tended to “communication”, like personal communication and real communication in English. Also, formulaic language teaching was implicit (in conversations with foreign teachers) and explicit (instructions in the classrooms and personal advice).

Additionally, the researcher also gathered the interviewees' comments on the learning context at the research site. On the one hand, based on the analysis of internal factors mentioned by the interviewees, it is noted that the interviewees had high expectations of their English language performance and did not want to show their shortcomings in English; as for English learning, they realised that lack of the immersion in the context and the accumulation in the language learning, as well as their relatively passive learning attitudes led to various challenges "bottleneck" in English language learning. On the other hand, external factors like shortages of equipment and facilities for English language learning, the urgency to improve the quality and quantity of teaching staff (which has already influenced syllabus design), some pressure from peers and low level of willingness to communicate in English, played a negative role in English learning at the research site.

Using interview research methods, the research managed to gather and analyse data based on a combined inductive and deductive approach to present the data extracted from the interviews in this section, with five main aspects, understanding of formulaic language, formulaic language learning, using formulaic language, formulaic language teaching, and learning context at the research site. Two rounds of interviews enabled the researcher to obtain in-depth data regarding formulaic language; also it allowed the researcher to investigate language phenomena that are not directly observable, in the case of learners' self- reported perceptions or attitudes (Mackey and Gass, 2005). Also there are some drawbacks within the research instrument itself, for example, the interviewees may portray themselves better than in reality. This could result in the halo effect which refers to what happens when interviewees pick up cues from the researcher related to what they think the researcher wants them to say, thus potentially influencing their responses (Mackey and Gass, 2005). Therefore, the results need to be compared with the results of the written texts analysis and interviews.

6.3 Triangulation of findings

As two of the four aspects of triangulation (triangulation of methods, investigators, data and theory), the triangulation of methods and data allow the researcher to get as full and as accurate a picture of the research problem as possible (Denzin, 1970). In this study, I used two research methods (small student written corpora and interviews) to collect and analyse two types of data: textual data and perceptual data, in order to answer the research question and depict the learning and use of formulaic language in Chinese university students.

Overall, the results analysed in the preceding sections corroborate across two types of data but collected using different method: textual data collected from the student written texts and perceptual data from the interviews.

Frist, the triangulation was realised through the perceptual and textual data. For the general recognition of formulaic language, from the data collected from three methods, most participants in the study showed their awareness of the existence of formulaic language in English learning and use, especially in written texts; however, their awareness of formulaic language in English-speaking remained unclear. Also, the interviewees showed that they paid more attention to fixedness as one of the most distinguished features of formulaic language than frequency, which can be seen clearly from the interview transcripts, and the student identified formulaic strings in their written texts. Also, interviewees could provide general features of formulaic language but not the specific features of each type, for instance, they recognised and identified formulaic language, and they used terms such as idioms and phrases, but they were not able to offer a description of the structure or functions of those terms. The one exceptional was the templates that they used in structuring expository or recount essays.

The collaborated part of the results above demonstrates the advantages of using triangulated methods to investigate the learning and use of formulaic language with two

types of data. As a compensating approach, some explorative questions were asked in the interviews, and these provided more detailed descriptions or explanations of certain issues, for instance, the students' perceptions of the structural and functional features of formulaic language (see section 6.1).

For the learning formulaic language, the data showed that participants were interested in learning phrases, which was consistent across the textual data collected by different methods. However, there seems to be a discrepancy between learners' needs and teaching in the classrooms in terms of phrases, as it was reported by the participants that the teaching of phrases in the classroom focused on the structural or formal aspects, rather than detailed explanations of the role or function of the phrases.

Second, the triangulation of findings was also achieved within the research instruments, interviews, and written corpora respectively. Regarding the use and learning of formulaic language, the results from the two rounds of interviews, show that there was no clear-cut distinction between learning and using for the participants. Mostly, these two aspects were intertwined with each other.

The data drawn from the two rounds of interviews also show that there is a shift in language teaching from middle schools and high schools where teachers emphasise memorizing formulaic language to students in the classroom, to the university English classroom where participants reported that their teachers tried to encourage them to learn and use English in a more general sense.

The triangulation of findings was also conducted in the written text analysis. After classifying the identified formulaic strings and clusters structurally and functionally, the researcher also compared the two types of formulaic language across two corpora. The shared formulaic strings and four-word clusters and core formulaic language are typical of the triangulated findings within the textual analysis.

6.4 Limitations of the findings

There are some limitations concerning data. There are two methods from both research strands for data collection, which involve including interviews and students' written texts. As discussed in Chapter 5, as qualitative research instruments, interviews have strengths as well as limitations (see section 5.10). For the strengths, for example, two research questions rely on the perceptual data that I mainly gathered from the interviews, it turns out that many compelling, detailed and even unexceptional issues emerged from these kinds of data. Also, in order to minimize the problems with the interviews, for instance, the "halo effect" (see section 5.8.2), the data and results were analysed individually and finally triangulated and synthesised in Chapter 7. This helps to increase the reliability and credibility of the research.

However, there are still some other problems with the data collection, taking interviews as an example. The timings of the interviews are different within the group, sometimes, the interviewees recognised the issues I asked about in the interviews, then they had much to say; sometimes, they may have never thought about the notion of formulaic language before, so it could a bit tricky for the students, in fact, they were not familiar with the topic. Therefore, they did not talk that much during the interviews, thus, they were not able to provide enough information regarding some more specific issues, for instance, the language learning strategies they used to learn specific types of formulaic language. Actually, based on the findings of this research, it is noted that students seem not to be bothered with the frequency of the formulaic language they use, instead, they paid more attention to how to express themselves in English more clearly, accurately, appropriately and fluently.

In addition, during the fieldwork, due to the situation at the research site, it was not possible for the researcher to control the length, choice of topics or the quality of the written texts. Therefore, the texts that I collected from the research site are shorter than expected, as they were set by teachers for different purposes, which limits the size of the learner corpora. It is also noted that during the process of transcribing data from

handwritten form to the MS word document, some texts were found to have been completed by cutting and pasting paragraphs from texts on the internet, as the purpose of the submission was for attendance checking after a holiday (see section 5.6.1). More rigid research agenda are called for in future research. It is necessary to take certain factors into account when investigating formulaic language through written texts, such as the length of the texts, the writing tasks, whether students can access reference material while writing, timed or untimed writing, and so forth. In addition, the process of transforming the handwriting into electronic documents to compile the student written corpora is extremely time-consuming and laborious.

Due to the research focus of this study, there are some questions which remain unanswered. There are some interesting cases of formulaic language use, for example, compound words and conjunction words. In the student identified formulaic strings, the initial list contains individual words, including hyphenated compounds, conjunction words, and words containing inflectional or deviational parts; whereas in the interviews, compounds were also mentioned by the interviewees as one type formulaic language in their perceptions. In a previous study on formulaic language, Wood (2015) considers compound words as special cases in formulaic language study as they indicate strong lexicalisation. Studies conducted by Chen and Baker (2010) and Leedham (2011), found that L2 learners prefer certain conjunction words and those conjunction words were regarded as chunks in corpus research. However, due to the scope of this research, these cases were ruled out for further analysis, as they are not multi-word combinations.

In conclusion, while there are some limitations to this research, still, this explorative study on formulaic language used by Chinese university students using mixed-method research could provide descriptions as well as some explanations of formulaic language use from the learners' and users' perspectives. At the same time, it highlights some specific issues in English formulaic language learning and teaching in EFL contexts. This study aims to contribute to a better understanding of the essence of formulaic language,

and to provide some recommendations for future research on formulaic language as well as the teaching of formulaic language.

6.5 Summary

In this chapter, first, the nature of each research question was re-clarified. The results and findings from each method (cf. interviews and written texts) were triangulated and synthesised under the research questions. The main findings are concerned with, on the one hand, the description of two types of formulaic language (formulaic strings and four-word clusters) based on the textual analysis, in relation to quantity, structural, functional, distributional relationships, and group differences. Two new concepts were introduced in the analysis: shared formulaic language and core formulaic language. On the other hand, the analysis of perceptual data leads to findings regarding the students' perceptions of formulaic language, their learning and use of formulaic language, and their comments on the teaching of formulaic language. Apart from this, the limitations of the research and some recommendations for the future research are also extensively discussed.

In general, through two identification approaches, that is, the speaker-internal (psycholinguistic salient formulaic strings) and the speaker-external (four-word clusters extracted by corpus linguistics approach) approaches, formulaic language in the students' English written texts was identified and compared. In terms of the corpus linguistics approach, the distributions of broad structures and functions are mostly in line with the results from previous research. The comparison between these two types of formulaic language introduces the concept of core formulaic language, which refers to the shared or overlapped formulaic strings and clusters in one corpus. By comparing formulaic language use across the two year groups, the concepts of shared formulaic strings and clusters were developed. The shared formulaic strings and clusters in this study refer to the shared or overlapped formulaic strings and clusters in one corpus. This indicates that in spite of the differences across the year groups, there are still some shared aspects.

In order to provide insightful information for two sets of research questions (RQ 2, 2a and RQ 3), perceptual data were collected from the interviews to report students' perceptions of formulaic language and related issues. The findings indicate that there are noticeable gaps in the perception of formulaic language between language users (students) and researchers. For instance, certain features of formulaic language were reported by the students in the interviews, but have been ignored in the research in this area, such as the analytic roles of L1, and the literate effect produced by formulaic language. It is also noted that students noticed more about the fixedness of formulaic language than the frequency of formulaic language use. The reason for this could be the different views of language between held by students and researchers. Thus, students, as language users, pay attention to specific situations involving language use, whereas researchers tend to find general rules or patterns from massive language production (see section 7.2). This may explain why students had different attitudes towards the repeated use of formulaic language in their own texts and in their peers' texts.

The last research question (RQ 3), aims to explore related issues concerning formulaic language teaching and learning at the research site. Through the discussion, it was found that formulaic language was not clearly defined by the researcher, or taught by the teachers. Therefore, the students were not able to provide anything but a relatively superficial understanding of this language phenomenon. More research is needed to build up the common ground in understanding formulaic language among researchers, students, and teachers. In addition, it also reflects some problems with English language teaching in the university, for instance, the teaching approaches were still dominated by traditional audiolingualism and only focused on the form without functional teaching at all. In other words, there is a lack of effective communicative exchange between teachers and students.

Overall, this study provides insights into formulaic language and related issues from the student perspective, which lays the foundations for future research on this strand by providing well-rounded recommendations for future work (see section 7.4). However,

the positioning of emerging English formulaic language with specific cultural and social meanings needs to be examined in further research, especially in the EFL or intercultural context. In the next chapter, based on the findings and discussion in the preceding chapters, the conclusion and some implications for formulaic language teaching will be introduced.

This chapter presented the analyses and results drawn from two types of data (textual and perceptual data), which were collected using two methods: interviews, and students' written texts. In each section, an overview of the data and the analytic steps was provided first. The analysis follows the analytic procedures proposed in Chapter 5. First, the results and findings of the textual analysis were presented. Two approaches were used to identify formulaic language in learners' written texts: speaker-internal (psycholinguistic salient formulaic strings) and speaker-external (four-word clusters extracted by corpus linguistics approach). In the analysis of the interviews, both bottom-up and top-down approaches were employed, in order to keep the analysis within the research focus, as well as to capture unexpected issues reported by the interviewees. To compare the results from different methods, the interview data is presented in four themes, these are: understanding formulaic language, learning formulaic language, using formulaic language, and comments on formulaic language teaching. The synthesis of results and a detailed discussion regarding the research questions will be presented in the next chapter.

Chapter 7 Discussion

The preceding chapter has presented the analysis and results regarding the research questions based on each identification approach. This chapter will focus on the synthesis and discussion the findings related to each research question (see section 5.1 for research questions). First, it will answer the first main research question "to what extent do Chinese students use formulaic language in their own texts", by examining and discussing the results in terms of the main structural, functional and distributional relationships between the categories, as well as exploring the group differences in formulaic language use (section 7.1). Followed by this, the second research question (2. What do Chinese university students perceive formulaic language to be?) and its sub-question (2a. To what extent is this perception different in Year 1 and Year 3 students?) is answered in section 7.2, which is based on the findings from the interviews and textual analysis. Section 7.3 will present the students' perception of the learning, use and teaching of formulaic language in the research context by discussing the findings from the interview analysis (3. How do Chinese university students perceive the learning, use and teaching of formulaic language?).

By doing so, it is hoped to present a comprehensive and accurate picture of formulaic language use by the Chinese university students. This study ultimately aims to provide some implications of teaching formulaic language in higher education in the EFL context. The chapter will end by discussing some of the limitations of the current research project before arriving at the conclusion chapter.

7.1 Formulaic language in Chinese university students' written texts

The first research question asked "To what extent do Chinese university students use formulaic language in their written English?" Additionally, three sub-questions were included addressing the main structures and discourse functions of the formulaic language used in the written texts, as well as asking for an exploration of the group

differences in formulaic language used in the Year 1 and the Year 3 students' written texts.

7.1.1 The main structures of formulaic language used in the written texts.

This study examines two types of formulaic language, i.e. formulaic strings that were identified by the students according to the psycholinguistic salience, and four-word clusters, which were extracted based on frequency account with the corpus linguistics approach.

As for the formulaic strings, they are a type of formulaic language and they were identified by a speaker-internal approach. Thus, students marked those language expressions that they thought they had learned and used as a whole once they finished the written texts. In other words, the formulaic strings were psycholinguistic salient for the language user. The results of the identification show that the formulaic strings range from individual words to templates (see section 5.8.1). I also observe that most strings are multi-word, semantic and functional units but they are not necessary complete in structure, for example, *I firmly believe that we*.... They can be continuous, for example, *as well as*, *more or less*, *the dragon boat festival*, or discontinuous, like *Only when…can we*, *first…second…last but not least*, *bring…up*. However, due to the focus of this study being on multi-word units, individual words were excluded from the analysis, which mainly comprised the deviational or inflectional changes of a word, conjunction words or hyphenated compound words.

Four-word clusters are the other type of formulaic language investigated in this study, and these were identified by the speaker-external approach, i.e. a corpus linguistics approach based on frequency of occurrence account. A series of selective criteria were set up to finalise the list for analysis, including the length of the cluster (which is four-word), the normalised threshold frequency, and text dispersion. In addition to this, all the clusters are continuous. In line with previous research on lexical bundles, the clusters mostly have incomplete structures and usually stretch between two structures.

Both strings and clusters were syntactically classified on the basis of the first part of the item, which is also the usual method adopted in structural classification.

Overall, there are three broad structural categories in the identified formulaic language in this study and they are phrasal, clausal and other. It is noted that this classification aims to provide a descriptive framework to depict formulaic language use in student English written texts, rather than a definitive criterion to judge if an expression is formulaic. Both the formulaic strings identified by the students and the four-word clusters extracted from the corpora follow a general pattern, that is, phrasal structural is the most frequently identified category, followed by the clausal structural and the other structures is the least frequently identified category. Also, the proportions of each broad structural category in the strings and clusters are very similar (see Tables 6.1 and 6.2).

As the most identified structural category, formulaic language (i.e. formulaic strings and four-word clusters) with phrasal structures takes up the majority. There are four sub-structures under this broad category, which are VP-, NP-, PP-based and Adj. P/ Adv. P. Among these four sub-structures, the most identified is the VP-based structure, whereas the least identified is the Adj. P/ Adv. P structure.

The VP-based structure is divided into a further five categories: 1) copula be + NP/Adj. P; 2) Anticipatory it+ VP/Adj. P+ complement clause; 3) passive VP (fragment); 4) pronoun/ NP+ BE; 5) active VP (fragment). The Active VP (fragment) and pronoun/ NP+ BE are the two most popular structures in the identified strings and clusters. The phrasal verbs and There be are the two most identified structures under the VP structural category (see section 6.1.1.1). No formulaic language with anticipatory it structures was found in the clusters. Also, the strings and clusters with passive VP (fragment) did not take up as much as previous studies indicated, despite being considered as a significant feature in academic writing (Biber, Conrad, and Cortes, 2004; Chen and Baker, 2010; Hyland, 2008b; Leedham,2001; Salazar, 2011). The reason for this could be the nature of the writing tasks; these were more subjective than most academic writing tasks, as most of them were based on the writer's personal experience and general knowledge of the real

world, without the textual plurality (see section 6.1.2). Regarding the NP- and PP-based structures, more formulaic strings with PP-based structures were identified than clusters; whereas more clusters with NP-based structures were identified than strings. It can be assumed that this was caused by the methodological option "stop at punctuation" in the Wordsmith tool 6.0 when extracting four-word clusters (which means any clusters near a punctuation mark were excluded), whereas most formulaic language with a PP-based structure is located near punctuation (see section 6.4).

There are six sub-categories found under the second broad structural category, clausal. These are: (verb/adj. +) to- clause, as-clause, if-clause. (NP+) (Verb) that-clause, 1st/2nd person pronoun +dependent clause, and WH-clause. For both the formulaic strings and four-word clusters, the most identified structures in the clausal category are WH-clause and to-clause. Whereas the least identified structures are that-clause and if-clause. Moreover, no strings with that-clause were identified. It can be assumed that students identified most strings with clause led by the conjunction word That as a part of the other strings, for example, *I believe that*, *I hold the view that*. In addition to this, there is a difference in the identified formulaic strings and clusters with WH-clause structure. Thus, most identified formulaic strings with WH-clause structure are exclamation sentences, whereas most clusters with this structure are dependent clause constructions.

The last structural category, i.e. 'other' comprised four sub-categories, i.e. ill-formed, formulae, templates and other expressions. There are no clusters identified as ill-formed or formulae. Ill-formed formulaic language usually has an irregular syntactic structure, thus it is also called irregularity in form and has been the focus of previous research on formulaic language or idioms for a long time (Moon, 1998). The findings from this research prove that ill-formed or irregular items do not appear frequently enough to be identified as formulaic language in the corpus linguistics approach.

For the formulae, there are several types of formulae used by students that could be intelligible in a wider community: 1) original formulae were used, for example, "*no pain,*

no gain"; 2) structure-kept expressions, for example, "*where there is [a]river, there is a city"*, which actually comes from the proverb "*where there is a will, there is a way"*, in which the structure remains while the content is changed; 3) content kept only expressions where only the literal meaning is kept, for instance, formulae used by students "*The rose is in her hands, the flavour in mine*", or "*rose to lover, smell in hands*" is originally from the sentence "*the fragrance always stays in the hand that gives the rose*", which comes from a poem. However, this type of expressions is mostly ungrammatical and features Chinese English or Chinglish, which is often an actual word-for-word translation from Chinese to English that is rejected by the Standard English norm (see Wang, 1997; Kirkpatrick and Xu, 2002).

However, there are other formulae which can be very difficult for those outside students' socio-cultural background to understand. For example, "*where there is a book, there is gold*". The structure of this sentence comes from the proverb "*where there is a will, there is a way*", whereas the content is translated from ancient Chinese literature "*书中自有黄金屋 (shu zhong zi you huang jin wu)*", which means reading can make one change his/her life into a wealthy one.

The last sub-category "other expressions" includes all the fragmental and meaningless expressions. There are more clusters than strings identified in this category. This is probably because of the small size of both corpora, and because the extracted clusters show a more fragmental tendency.

In conclusion, this section has mainly discussed the structural patterns of two types of formulaic language, i.e. formulaic strings identified by the speaker-internal approach, and four-word clusters identified by the speaker-external approach. Based on the analysis and findings, I would argue that in terms of structure, the use of phrases is very pervasive, which is especially true of VPs (phrasal verbs and There be sentences), NPs and prepositional phrases. At the same time, formulaic language with clausal structures is also frequently used in the student texts, especially the WH-clause and to-clause. However, in student perceptions, formulaic language can range from the level of

individual words, for example, derivations or inflections of words, conjunction words and hyphenated compounds to the phrases, clauses and sentences (for instance, formulae), up to the macro-level of discourse categoriser: templates, which are used to structure the whole paragraph or text. This finding is in line with Hüttner's (2007, 2008) study, which found that formulaic language, for students, is not limited to the micro-level, such as phrasal studies or multi-word unit studies, but expands to the macro-level, even in a generic sense, so that they have specific templates to help structure specific English written texts overall. In addition, from the analysis of formulae, it can be seen that students' L1 plays a role in formulaic language learning and use, which is especially apparent in the fact that the translation of formulaic language swung between English and Chinese, which does not happen with monolingual speakers. However, this can cause some problems with intelligibility in the communication, especially when the students change the content of the formulae and try to express their own meaning.

7.1.2 The main discourse functions of formulaic language in written texts

Based on previous research (Biber and Barbieri, 2007; Hyland, 2008a; Leedham, 2011; Salazar, 2011), a typology of discourse function of formulaic language was proposed (see section 5.8 and section 6.1.2). Following Halliday (1978) and Martin (2009), a natural relation is posited between the organisation of language and the organisation of social context, built up around the notion of three types of meaning, i.e. textual, interpersonal, and ideational. The textual functions are similar to Halliday's textual metafunctions and Biber's text organisers in that they are connected with the overall organisation of the text; the interpersonal function is used to provide a structure for interpreting a following proposition; last functional category ideation, helps writers to structure their activities and experiences of the real world (Hyland, 2008a) (also see sections 5.8 and 6.1.2).

The multi-functionality of formulaic language cannot be overlooked, as some formulaic strings and four-word clusters serve more than one function (see section 5.8.1). However, in this study I adopted a practicable solution to the functional classification,

which is based on the primary or the most common function that a formulaic expression serves.

Overall, the formulaic strings and four-word clusters all follow a general pattern in their primary or most common function: the most identified is the ideational function, followed by the textual function, with the interpersonal function as the least frequently identified.

As the most frequently identified functional category, the ideational function has four sub-categories: quantification, location, description and action. Among these sub-functional categories, formulaic strings and four-word clusters with a description and action function take up the largest proportions. However, there are more formulaic strings identified with an action function than a description function, whereas there are more clusters identified with an action function than a description function. In addition to this, there were no substantial differences in the quantity of identified formulaic strings and clusters with quantification and location functions.

The textual functional category comprises the most sub-functional categories among these three broad functions, including resultative, structuring, generalisation, transition, exemplification, and framing functions. I added exemplification and generalisation to the functional categories due the nature of the writing tasks in this study, which is very subjective, and so students need to use examples from their personal experience or general knowledge of the real world to exposit their opinions on certain topics (see chapter 5). The findings show that transition is the most identified textual function in both strings and clusters, whereas generalisation is the least identified function in both types of formulaic language. In addition, structuring and exemplification were not identified in the four-word clusters.

The overall pattern under the last broad functional category interpersonal function is that more strings and clusters were identified with a stance function than an

engagement function. This finding is in accordance with previous research (Hyland, 2008a).

To sum up, this section has discussed the functional categories of the identified formulaic strings and four-word clusters. The results from the analysis of both types of formulaic language show a shared tendency for the ideational function to be the most frequently identified function, followed by the textual function.

The ideational function is the most frequently identified function, which students often used to express their meanings and to refer real world experience and activities, especially with the description and action function. From the six sub-functions of the textual one, transition was found to be the most common, which makes a text coherent and cohesive. The least frequently identified function was the interpersonal function, where there is more formulaic language with a stance function to express the writer's attitude and evaluation than the engagement function, which addresses the reader directly. This also reflects the nature of the texts, as the tasks required were subjective, and based on personal experience or general knowledge of the real world.

In order to gain further insights into the structure and function of formulaic language used by the students in their English written texts, in the next section, the relationship between the structural and functional categories of formulaic language identified by both approaches will be discussed.

7.1.3 The relationship between the structural and functional categories

In order to provide sufficient evidence to answer research question 1c, this section is going to firstly look at the relationship between the distribution of structural and functional categories in the formulaic strings and four-word clusters. It will then turn to the core formulaic language, which consists of the shared formulaic strings and clusters in one corpus (see section 6.1.5), to discuss the relationship between the structure and function of formulaic language in general. It is hoped that this will help to pin down the

positioning of formulaic language in this research and further contribute to the description of formulaic language use in the research context.

As discussed in section 6.1.6 and shown in Figures 6.17, 6.18, 6.19 and 6.20) the most identified strings and cluster with VP-or NP-based structures usually serve an ideational function; most clausal structured strings and clusters perform an interpersonal function in the written texts; strings and clusters with PP-based or under the third broad category "Other", are usually associated with an ideational or textual function. In other words, the ideational function is mainly realised by the VP-based, PP-based, NP-based and other structures; the interpersonal function is mainly realised by formulaic language with a clausal structure, whereas the textual function is realised by PP-based and other structures.

This can be compared with previous research on lexical bundles, for example, a study by Biber and his colleagues (2004). They suggest that there is a strong relationship between the structural and functional categories for lexical bundles. Their research indicates that most stance bundles are composed of clause fragments, while most referential (ideational) bundles are composed of NPs or PPs fragments. These results are in line with the result from the current study. At the same time, they also suggest that these patterns are strongly associated with register, as conversation and academic prose constitute the two extremes of oral and written registers and at the oral extreme, VP-based and dependent clause lexical bundles are mostly used for stance functions. Academic prose, at the literate extreme, uses mostly NP/PP based lexical bundles for the referential function. In the current study, however, the most frequently extracted formulaic language is ideational strings and clusters with VP-or NP-based structures. This finding indicates that the students' written texts in this study are a mixture of oral and written features, which contradicts much research on academic writing. The reasons for this could be the nature of the writing tasks, compared to traditional academic writing, as they are more subjective and do not involve of textual plurality; and/or because the

students did not receive clear instructions or have enough awareness of difference between the oral and written language.

In addition, Biber, Conrad and Cortes (2004) interpret the reliance on NP/PP-based bundles as evidence that lexical bundles are stored as unanalysed multi-word chunks, rather than as instances of productive grammar construction and argue that referential bundles composed of NP and PP are stored and used as single units, without reference to their structural correlates. However, in this study, the students were asked to mark the formulaic language that they used in their own English written texts, thus, it is necessary to examine the core formulaic language, and therefore gain insight into the relationship between the structural and functional categories.

In section 6.1.5 I outline that both formulaic strings and four-word clusters can be incomplete in their syntactic structures and that they can be classified in the same structural and functional categories, for example, *one of the most* (cluster) and *one of the* (string). However, the structural boundaries can be very different. For instance, the string *I firmly believe that we…* has an incomplete structure, yet it still makes sense to the language users in that the fragmental part *that we* can be considered as a grammatical reminder to help students figure out the grammatical construct, which is a complement clause, in the following part to finish the sentence correctly. Therefore, in the language users' psycholinguistic status, a formulaic string can be incomplete in syntactic structure, as can clusters, however, the strings make sense to the language users, by serving as a semantic unit, like *try our best to*; or as a grammar reminder to help achieve correctness.

This finding corresponds with previous research on formulaic language learning (Mitchell and Martin, 1997; Myles, Hooper, and Mitchell, 1998; Weinert, 1995), in which the researchers showed that formulaic language plays a central role in language learning, feeding directly into the learning mechanism, through a process of gradual analysis and unpacking the chunks eventually. However, for learners at the early stages, those

broken-down chunks may drive the construction of the L2 grammar forward, and this seems to be the case with the learners in this research.

Second, despite the blurred boundaries between the forms of formulaic language, the function remains relatively stable, for example, the core formulaic language pair: the cluster *on the other hand* (cluster) and the string *on the one hand…on the other hand*. They have different structural classifications, the cluster *on the other hand* was identified with PP-based phrasal structure, whereas the string, *on the one hand…on the other hand…* was classified as a template under the "other" category. However, their function remains the same and they both serve as transitional signals in the textual function. This finding is in line with the study by Durrant and Matthrews-Aydınlı (2011), in which the researchers used a function-first approach to identifying formulaic language in academic writing. Also based on the analysis and finding in section 6.1.5, there is a clear trend in the function distribution of shared and core formulaic language, which is that textual is the most identified function, followed by the ideational and interpersonal functions respectively, whereas the structural distribution of the shared and core formulaic language shows a more chaotic picture.

In conclusion, this section discussed the relationship between the structural and functional categories by examining the distributional relationship between structure and function and it was noted that the findings from this research are in line with the previous research (Biber, Conrad and Cortes YEAR). However, it also indicated that the formulaic language used by the Chinese students in this study is a mixture of oral and written registers, as both VP-based and NP-based structures are mostly identified as showing the ideational function. On the other hand, by investigating the core formulaic language, i.e. shared and overlapping formulaic strings and clusters in one corpus, I could establish that both formulaic strings and clusters can be incomplete in syntactic structure. However, the formulaic strings still make sense to the language user, either because they have semantic meaning or because they serve as a reminder for the grammar. This finding is in line with the study by Myles et al. (1998), which concludes

that formulaic language (chunks, in their terms) provides a set of language samples, which act as models for constructing grammar. Moreover, it is also found that the structure boundaries can be blurred by different language users' psycholinguistic salient status or through different corpus extraction approaches, but that nevertheless the function of the core formulaic language pairs remains the same. This finding can shed some light on the nature of formulaic language in future studies.

7.1.4 Group differences: formulaic language used by the Year 1 and Year 3 students

The written corpora used in this study consist of the English written texts from two groups of students from Year 1 and Year 3. Given that the students in the senior years had been studying English as their major for much longer than the Year 1 students, who had just graduated from high school, the English language proficiency and expertise in English writing of the Year 3 students is more advanced. It is assumed that there is a difference across these two groups in terms of formulaic language use as they are writers with different English language proficiency levels and writing expertise (see Biber, Conrad and Cortes, 2004; Hyland, 2008a; b Myles and Cordier, 2016). Next, the two groups will be compared in terms of the quantity, structure and function of formulaic language, and in terms of the shared and core formulaic language.

First, as for the quantity of formulaic language used in the written texts, from Table 5.8 (section 5.8.1 phase 2), it can be seen that the Year 3 students tend to use more and longer four-word clusters, as there are more three-, four- and five-word clusters extracted from the Year 3 corpus than the Year 1 corpus. After careful calculation, as discussed in section 6.1.4.1, overall, the total number of formulaic string and four-word clusters is greater identified in the Year 3 corpus than in the Year 1 corpus. The average numbers of formulaic strings and four-words went against this trend (see table 6.20 in 6.1.4.1): the average numbers of identified strings in the Year 1 and Year 3 corpora are 1.12 and 0.85 respectively, whereas the clusters are 0.84 in the Year1 corpus and 1.07 in the Year 3 corpus. In general, it is difficult to say which groups used more formulaic language in the written texts. This finding contradicts the results from previous research

(Hyland, 2008a), that argue that the number of lexical clusters will decrease as language proficiency or writing expertise increases. The factors that result in this contradiction could be the differences in terms of methodology (see section 6.4).

Second, the findings regarding the structural classification indicate that for both year groups, formulaic language (i.e. formulaic strings and four-word clusters) with a phrasal structure is the most identified, followed by formulaic language with clausal structure; the least identified structure is the "other" category. Also, the proportions of identified formulaic strings and four-word clusters are very similar (see Tables 6.1 and 6.2). There is no difference between the two groups in terms of the three broad structures.

However, there are some differences in the sub-structural categories across the year groups. Under the phrasal structures, the Year 3 group tended to use more formulaic language with VP-based and Adj. P/Adv. P structure. The possible explanations for this could be that the Year 3 students in general have a larger vocabulary size or are better at describing events or actions in detail than the Year 1 students. Among the clausal structures, the Year 1 students used more formulaic language with the WH-clause (mainly exclamation sentence led by WH-word) and 1st/2nd personal pronoun + dependent clauses to express their stance. In the final structural category "other structure" the Year 3 generally used more formulaic language than the Year 1 students under each sub-category.

Next, when it comes to the functional categories, the overall pattern shows that both groups shared a trend in the functional aspects of formulaic language, where the ideational comes as the most common function, followed by the textual function and then the interpersonal function. However, in the proportions of formulaic strings and four-word clusters identified with three broad functions, it can be seen that in the Year 3 corpus, the ideational function increased by about 20%, which ranges from 60% to 80%; whereas the textual and interpersonal functions decreased by about 10% respectively (see Table 6.1). This indicates that the more experienced language users tend to use more formulaic language to make meaning or convey information in their texts.

The sub-categories under each broad function are complex. With regard to the textual function, overall there are more formulaic strings and clusters identified in the Year 3 corpus than in the Year 1 corpus. However, in the sub-function categories: exemplification, generalisation and structuring, more formulaic language is used by the Year 1 students than the Year 3 students. The only convergence between both groups is that the transition function takes up the largest proportion of the formulaic language identified in the two corpora. As for the interpersonal function, the two groups still share the general trend that there is more formulaic language with a stance function than with an engagement function, however, Year 1 students tend to use more formulaic language with an interpersonal function than the Year 3 students. For the final category, ideational function, there is less formulaic language identified in the Year 1 corpus than in the Year 3 corpus. Again, it can be assumed that the Year 3 students, as the more experienced writers and more proficient language users, have a richer repertoire to express themselves in writing.

It is also interesting to take a look at the shared formulaic language identified in this study. As shared formulaic strings and shared four-word clusters in the two corpora can tell us to what extent the formulaic language produced by language users with different language proficiency levels and writing expertise overlap. The findings show that the quantity of shared formulaic language is relatively stable regardless of the language proficiency level or the writing expertise of the language user. As there are 21 pairs of shared strings in the Year 1 and Year 3 corpora, and 20 pairs of shared clusters in the two corpora, the number of shared strings and shared clusters are quite similar.

Overall, in this section, the discussion regarding the first set of research questions (1a to 1d) is presented. The most popular structures of formulaic language identified by the students' perception and corpus linguistics in this study were phrasal, clausal and other structures (see section 7.1.1). The main discourse functions of the identified formulaic language were ideational, textual and interphone (see section 7.1.2). Also, this section talked about the relationship between the structural and functional categories (see

section 7.1.3). Last, the assumed group differences in terms of language proficiency and English writing expertise in formulaic language use were also explored (see section 7.1.4).

To sum up, the main structural patterns of formulaic language used in Chinese university students' written texts are extremely varied: they range from individual words, to the micro-level, such as phrases, clauses and sentences, like formulae, up to the macro-level, for example, templates for organising the whole text. Considering the focus of this research, the individual words were ruled out and the formulaic language in this study mainly covers three broad categories: phrasal, clausal and other. Formulaic language with a phrasal structure is the most identified, especially VP-based (mostly active phrasal verbs) and there be sentences; followed by formulaic language with clausal structure, in which the WH-clause and to-clause are very popular. The "other" category is the least identified, and its sub-categories ill-formed and formulae were not identified in the clusters, whereas the formulae strings can cause intelligibility problems in the communication.

As for the main discourse functions, based on previous corpus linguistic research, all the identified items of formulaic language can be categorised into three main functions, including ideational, textual and interpersonal. Overall, the most identified function is ideation, followed by the textual and interpersonal functions respectively. Among the sub-categories of each broad function, transition under the textual function, stance under the interpersonal function, and action and description under the ideational function are the most identified categories

This section has also discussed the relationship between the structure and function of formulaic language by examining the distributional relationship between the structural and functional categories (section 7.1.3) and then the core formulaic language, i.e., formulaic strings and cluster shared or overlapped in one corpus (section 7.1.4). For the distributional relationship, the results show that the ideational function is mainly realised by VP-based, PP-based, NP-based and other structures; the interpersonal

function is mainly realised by formulaic language with a clausal structure, whereas the textual function is realised by PP-based and other structures. By analysing the core formulaic language, I found that both formulaic strings and clusters can be incomplete in terms of syntactic structure, but formulaic strings with an incomplete structure, still make sense to the language user, either because they have semantic meaning or because they serve as a grammatical reminder. Moreover, I found that the structural boundaries can be blurred, by different language users' psycholinguistic salient status or different corpus extraction approaches; however, the function of the core formulaic language pairs remains the same.

The last part of this section has discussed the group differences in formulaic language use. First, I argue that it is difficult to say which group used more formulaic language by using different means of calculation, including the total number of different types (not tokens) of formulaic strings and four-word clusters, as well as the average number of different types of formulaic language in each text. Also, through the analysis of shared formulaic strings and clusters in two corpora, I could show that regardless of language proficiency and writing expertise, the amount of shared strings (21 pairs) and shared clusters (20 pairs) in the Year 1 and Year 3 group texts are quite similar. As for the structural categories, both groups follow the general trend that the phrasal structures are the most identified, the clausal structures come as the second, and the other structures are the least identified. For the functional aspects, comparing the proportions of formulaic strings and clusters, shows that in the Year 3 corpus, the ideational function increased by about 20%, ranging from 60% to 80%; whereas the textual and interpersonal functions decreased by about 10% respectively (see Table 6.11). This seems to suggest that more experienced language users tend to use more formulaic language to make meanings or convey information in their texts.

7.2 Students' perceptions of formulaic language

As stated at the beginning of this chapter, the preceding section aims to answer the first set of research questions (1a to 1d), which are based on the identification of formulaic

language by students' perceptions and corpus-based analysis. In the last section, some findings regarding the first set of research questions and its sub questions were discussed, while in this section, the focus of the discussion will exclusively be on the perceptual data collected from the interviews.

This section aims to provide the answer to the second set of research questions and sub questions (2 and 2a), as well as synthesising the results from different research methods into a coherent picture to explore the students' perceptions of formulaic language. It is also a turning point that directs the current study from language description towards language explanation (see section 3.1.3).

7.2.1 General understanding of formulaic language

As discussed in the previous sections, students identified the formulaic strings based on their perception of the holistic units. It is noted that identified strings like phrases, clauses and sentences as formulaic language. Students, on one hand, recognised some words, such as derivations and inflections of words, and conjunction words as formulaic language; on the other hand, they commented that templates that help structure the whole paragraph or even the whole text, and it can be considered as formulaic language. This finding again reconfirmed that formulaic language, in the students' perceptions, exists across the continuum with micro- and macro-level ends (Figure 7.1).

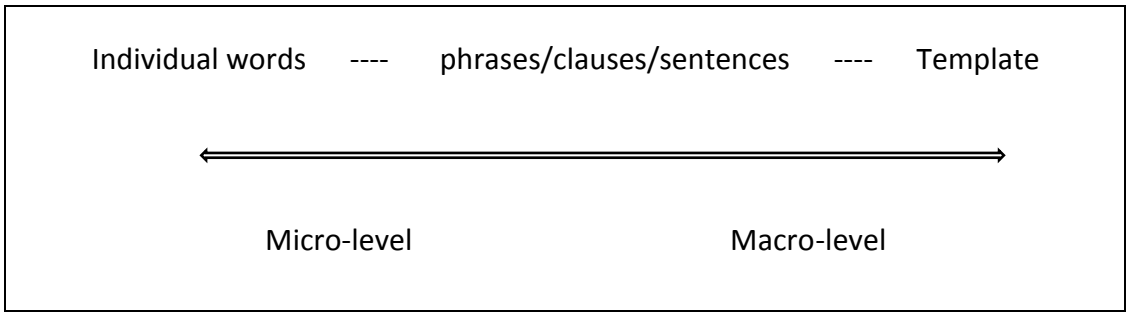

Figure 7.1: Continuum of the form of formulaic language

As shown in the figure above, formulaic language, in the students' perceptions is along a continuum ranging from micro-level to macro-level. For the micro-level end, formulaic language focuses on local lexico-grammar features, such as words, phrases, and some

clauses or sentences. As for the macro-level end, it is much more involved with the templates or frames that help to structure the text, for example, the students used *first…second… third…* to arrange their standpoints on expeditionary writing.

As discussed in section 2.3, two main aspects have been considered in previous research: form and function, which can help us to understand, identify and classify formulaic language. In the interviews, students talked about their understanding of formulaic language, especially in terms of the form and function (see section 6.2.3.2).

In terms of the formal aspects of formulaic language, interviewees from both groups mentioned fixedness as a feature of formulaic language, and used writing templates as an example. Moreover, interviewees generally considered "the good sentences and phrases" as formulaic language. Only two out of 12 interviewees claimed that high frequency was a feature of formulaic language. It can therefore be concluded that for the students, formulaic language is more closely associated with fixedness than frequency.

As for the function of formulaic language, students gave their own opinions on the role of formulaic language in English (see section 6.2.2). As reported by the students, there are certain aspects of English use that benefit from formulaic language: firstly, formulaic language was regarded as a "crutch", ensuring correctness and accuracy in English use for the students; secondly, it was convenient and handy to use ready-to-use formulaic language when needed, as it could be easily and quickly retrieved from memory, which saved the time and effort of having to construct new language every time. Thirdly, students claimed that formulaic language is sometimes a great addition to English use, as it can be used for producing funny or humorous effects. However, some students viewed formulaic language negatively, seeing it as too generic and rigid in form, and noted that using formulaic language put students at risk of making errors.

In this section, a discussion of the general understanding of formulaic language was presented, focusing on perceptions of the formal and functional aspects of formulaic

language. It is worth noting that students acknowledged the existence of formulaic language in English, and had a certain awareness of the form and function of English formulaic language in a general sense. Some distinct characteristics of formulaic language, which were discovered in previous research, were also recognised by the students in this study, for example, fixedness and functions of formulaic language in English use. Nevertheless, there are certain research-recognised features of formulaic language which were not acknowledged by the students, for instance, high frequency, which was one of the main criteria in formulaic language identification based on the corpus-linguistics approach. In order to further learn about students' perceptions of formulaic language, the following section will discuss students' understanding of specific types of formulaic language.

7.2.2 Understanding of specific types of formulaic language

In the first round of interviews, words like fixed phrases, collocations, idioms and fixed expressions, phatic expressions (including proverbs and greetings), writing templates and compound words were mentioned by the interviewees when they referred to or gave examples of formulaic language according to their perceptions.

In the interviews, students gave certain descriptions of features of formulaic language based on their perceptions, with regard to form and function. Students were able to give relatively detailed descriptions of certain types of formulaic language with focus on form and function, for example, writing templates. Yet, they only gave rough descriptions of other types of formulaic language without concrete examples.

As reported by the interviewees from both groups, the formal features of writing templates, can be the outlines of three- or five-paragraph essays or the specific sentences of a paragraph with open slots; whereas the functional features of writing templates were heavily relied on by the students in their English writing, especially in the exams, as the templates were useful to organise the texts and structure the information. In this way, the interviews argued that writing templates can make

students' English writing texts clear, as well as being a great addition to their language (see section 6.2.3.2).

As for the idioms and fixed expressions, none of the interviewees from either group were able to clearly describe their formal or functional features. However, they still talked about certain characteristics of idioms and fixed expressions according to their understanding. Idioms and fixed expressions, as reported by the interviewees, are usually fixed, with cultural connotations and remain within sentence-length (see section 6.2.2). With regard to the function of idioms and fixed expressions, there seemed to be a consensus that idioms and fixed expressions play positive roles in English language use; moreover, these two types of formulaic language could contribute to fluency in English and be helpful in terms of communication with people from specific cultural backgrounds.

There was not much about the form of the phrases in the comments of the interviewees. Only two out of 12 interviewees gave some description regarding the form of phrases (see section 6.2.2). However, as for the functional aspects, phrases were reported as being widely used by the interviewees and were used practically to express meanings. Furthermore, phrases can be used in an innovative way and thus make students' language improved and more readable. In addition, the correctness of phrases was usually stressed in the exams.

In terms of phatic expressions, in the interviews, rather than giving detailed descriptions with a focus on the form-function, students provided an integrated understanding of towards phatic expressions. Register and cultural differences were mentioned by the interviewees when they talked about their understanding of phatic expressions. Colloquial use, for example, was recognised as one distinct feature of phatic expressions, in other words, phatic expressions were restricted to spoken language, as they made the language informal and not appropriate and English writing usually requires formal language. Furthermore, phatic expressions were reported to be used especially in

communication with foreigners, as the phatic expressions in Chinese are different from those in English (see section 6.2.3.2).

As the last type of formulaic language examined in this research, compound words, according to the interviewees, are the product of word formation that creates compound lexemes, usually with transparent meanings. The interviewees also reported that the use of compound words would be a great addition to language production, as people usually do not use them that much; and that they could be a means of word formations to enlarge vocabulary size (see section 6.2.3). Terminologies was added by the researchers, and constitute words and compound words that in specific contexts are given specific meanings. These may deviate from the meanings the same words have in other contexts and in everyday language, for example, Incoterms stands for international commercial terms, and this is mostly used in international business studies, rather than everyday language. The university from which the interviewees were selected focuses its teaching and research objectives on English for Automotive Trading and Business English and this involves specific English content-based courses on Business English and technical words relating to vehicles. In fact, Business English and English for Automotive Technology predominate in English language teaching and learning in that university (see Chapters 4 and 5).

Based on the data collected from the interviews, it is clear that in terms of specific types of formulaic language, students had a stronger awareness of phrases, writing templates, idioms and fixed expressions, than the other types of formulaic language named by the students. Among these types of formulaic language, students were clear about the form and function of the relatively fixed formulaic language, for example, writing templates. However, when it came to other less fixed formulaic language types, for instance, phrases, idioms and fixed expressions, the students were not able to give clear and concrete descriptions with a focus on form-function. In conclusion, the students' knowledge and understanding of specific types of formulaic language was still limited and superficial, as they were not able to give concrete examples for all specific types of

formulaic language. In order to gain an insight into the students' perceptions of formulaic language, the group differences regarding the perceptions of formulaic language will be discussed in the following section.

7.2.3 Differences in perceptions of formulaic language between Year 1 and Year 3 students

In this section, the conceptualisation of formulaic language based on the students' perceptions was further discussed. It started with presenting the students' general understanding of formulaic language with a focus on the form and function of formulaic language. Then, the discussion focused on the specific types of formulaic language mentioned by the students in the interviews, in terms of form and function, followed by a comparison between the two groups of students' perceptions of formulaic language. In this way, it becomes possible to bring the textual data and perceptual data together to conceptualise formulaic language from the students' point of view.

For the general understanding of formulaic language with respect to form and function, students from Year 3 showed a broader understanding of formulaic language and expanded the vision of formulaic language. They provided examples of proverbs (example 4), as well as mentioning terms such as slang, phatic expressions and greetings were considered to be possible forms of formulaic language (example 5); they also started to think about some formulaic language with an opaque meaning (example 6), which could lead to difficulties in comprehension (see Chapter 2).

There is no significant difference between the Year 1 and the Year 3 groups in terms of their perception of the function of formulaic language. For students of both groups, formulaic language plays both negative and positive roles in their English use, but in general, the positive roles of formulaic language outweigh the negative in their English use.

As for the perceptions of the specific types of formulaic language, generally, interviewees from the Year 3 group paid more attention to the function of formulaic

language than the specific descriptive forms of formulaic language. Also, the students from the senior years showed a clearer understanding of the terminologies, whereas the Year 1 students seemed still to be struggling to tell compound words and subject-specific terminologies apart.

Students showed their awareness and understanding of formulaic language as a universal phenomenon across languages, including English. They pointed out some characteristics, with respect to the structure and function of formulaic language, which are mentioned in previous studies, for example, the fixedness and ready-to-use items. Beside this, students also mentioned some terms that they thought constituted formulaic language, and among these terms phrases, idioms and fixed expressions, and writing templates were mentioned more often than others. Students did mention some difficulties that can occur in comprehension or communication due to cultural or linguistic factors. However, they did not specifically raise norms of formulaic language in English. Rather, more attention seemed to be attached to formulaic language use in comprehension or communication.

Nevertheless, certain issues need to be taken into further consideration. First, in terms of the formal and functional aspects of formulaic language, although the students had some views, and indeed, the Year 3 students showed a deeper and broader understanding of formulaic language, still, the students' comments were limited to a relatively shallow level. Without any concrete examples to support their ideas or opinions on their conceptualisation of formulaic language, the students just talked about some ambivalent features, or simply repeated their teachers' instruction about formulaic language in the classroom. Students, for example, knew what a phrase was, however, they were not able to give a clear definition or specification for a phrase.

Second, the students' perceptions of formulaic language did not always match the researcher's. In the study, for instance, some students mentioned that the use of formulaic language under certain circumstances can produce a humorous effect, which as a theme has long been absent in formulaic language and English language learning

and teaching research. Students paid more attention to the fixedness of formulaic language rather than the frequency; this is probably caused by the different views of language between students and researchers: researchers may have an overview of the rules underlying the language, whereas the students only care about formulaic language use in a single situation.

In the next section, the issues related to students' perceptions of formulaic language learning will be discussed to further learn about how formulaic language is conceptualised by the students.

7.3 Students' perceptions of the learning, use and teaching of formulaic language

In this section, the issues related to the students' learning of formulaic language, which constitutes research question 3, will be addressed based on the perceptual data collected through the interviews. The results will be a comparison across the data collected using the mentioned methods, on three sub-themes: formulaic language learning, using formulaic language and students' comments on the teaching of formulaic language. These three sub-themes emerged from the process of analysing the interview data by using a combination of bottom-up and top-down approaches applied to the interview coding.

7.3.1 Formulaic language learning

According to the information provided by the students, the following themes concerning formulaic language from the perspective of English language learning emerged from this study: 1) the roles of formulaic language in English learning, 2) the process, strategies and sources that students would refer to during learning, 3) the learning preferences for learning six types of formulaic language.

First of all, all the students were asked about their views on the role of formulaic language in English learning. In the interviews, for example, interviewees were asked if formulaic language was important in English language learning. Seven out of 12

interviewees confirmed that formulaic language was important in their English language learning.

In order to gain an insight into the roles of formulaic language in English language learning from the students' perspectives, some questions in the interviews asked students about the influences that formulaic language had their English. It probably tells us that the students focused more on their own language performance, in terms of fluency and testing, rather than on (native) norms in their language production.

As for the processing of English formulaic language learning, most students were asked to recount the learning process in the interviews. Interviewees shared the view that this was a linear process (see Figure 6.22 in section 6.2.2), which started with attention or noticing, and then memorisation, followed by the review or retainment of formulaic language. The process was completed with the retrieval or recall of the formulaic language at the time of use (see section 6.2.3). Nevertheless, the interviewees from the Year 3 group added more stages to the process of learning, that is, comprehension, accumulation and independent learning of formulaic language beyond the classroom were considered to be beneficial (see example 33).

With regard to formulaic language learning strategies (see section 6.2.2), students from both groups shared some learning strategies, such as memorisation, accumulation and communication. These strategies were mostly used in the explicit learning of formulaic language. However, students from the Year 3 group stressed the classification, comprehension and independent learning, which are likely to happen in implicit learning (see section 3.1.2). They showed a broader repertoire and knowledge of formulaic language learning strategies than the Year 1 students. This is in line with previous research on language learner strategies, which shows that more proficient language learners tend to use a greater number of more integrated strategies to learn a language (Schmitt, 2000).

For the learning sources, based on the information provided by the interviewees (section 6.2.3.3), the students from both groups had the same or similar resources for

formulaic language learning, such as movies, TV dramas, textbooks, or reading. Almost all of these resources were visual. However, the interviewees from the Year 1 group, referred to dictionaries or foreigners as language learning resources to confirm the correctness and accuracy of their language, whereas the Year 3 group perceived themselves as language users, for instance, communicating with foreigners in English, in other words, they tried to use English and referred to different available resources rather than trying to achieve correct or accurate language production. It seems that the interviewees from the Year 1 tend to learn anything in English; however, the interviewees from the Year 3 group, assess the resources and aim for original or authentic English, rather than learning formulaic language unselectively from learning sources and materials in English.

Therefore, for the learning sources of formulaic language, to the students in this study, mostly the learning happens in the classroom, in an instructional context; for the learning materials, mostly they learn formulaic language through the written form, or visual learning. However, the students seemed not to receive enough clear instruction on formulaic language use in different registers, for example, spoken or written. This may cause an inappropriate transfer between the spoken language, which is often learned from visual methods, like TV dramas or novels and then written texts.

In terms of the learning preferences for specific types of formulaic language, in the interviews, the students were asked to rank the following six types of formulaic language according to their learning preference: phrases, compound words and terminology, idioms and fixed expressions, writing templates and phatic expressions. The results show that among these six categories, phrases, idioms and fixed expressions rank as the top three of the list for interviewees from both groups. The interviewees from Year 1 give priority to learning writing templates and compound words, rather than those from Year 3 who would like to learn some formulaic language like phatic expressions and terminology.

The interviewees were encouraged to talk about their strategies for learning specific types of formulaic language, i.e. based on their experience rather than classifying those

strategies. Interviewees from both groups realised that these six kinds of formulaic language are context-dependent. However, not all interviewees were able to report specific strategies for each type of formulaic language learning, which could indicate that they were not aware of all strategies used in learning. This may also be caused by a lack of explicit teaching or instruction from their teachers.

There is still lack of common ground as to what constitutes formulaic language among students, teachers, and researchers. If we bring the textual data and interview data together on the specific types of formulaic language together, it is worth noting that there are some overlaps as well as discrepancies among the typologies. The types of formulaic language mentioned by the students during the interviews were writing templates, idioms and fixed expressions, phrases, phatic expressions, compound words and terminologies, whilst the structural categories that appeared in the textual analysis are based on previous research on the structure of formulaic language, including clausal, phrasal and other structures. Among these perceptual and textual data, it can be seen clearly that formulaic language with phrasal structures plays an important role in formulaic language learning and use for the students in this research. Similarly, templates (a sub-category under the other structure category) and idioms (similar to the ill-formed and formulae under the "other" structural category) are another two types of formulaic language overlapped in the perceptual and textual data. It is also worth mentioning that compounds were identified as formulaic language in the interviews and formulaic strings. Due to the consideration of methodology, as one type of individual word, compounds were excluded from the list for analysis. However, compounds, in formulaic language study are regarded as special cases. Therefore, these structural types do not always overlap, for example, terminologies were mentioned by the interviewees a few times, but no terminologies were found in the student identified formulaic strings; on the other hand, there is a structural category named "words" in the textual data, where there are some morphological changes (inflection and inflection) with the words and conjunction words were identified by the students as formulaic strings in the texts but these were mentioned in the interviews at all (see section 6.1.3 and 7.1.1).

Based on the discussion of formulaic language learning in this section, it can be concluded that, firstly, not all the students in this research appreciated the role of formulaic language in English language learning. This indicates that there is not enough awareness of formulaic language amongst the students.

Secondly, the general learning process of formulaic language was shared by the students from both groups, which starts with attention or noticing, and then memorisation, followed by the review or retainment of formulaic language, and ends with the retrieval or recall of formulaic language at the time of use.

The main learning strategies reported by the students are memorisation, accumulating items and communication. The Year 3 students showed that they used more extensive strategies, such as comprehension, classification and independent learning, which may indicate that more implicit formulaic language learning happened among these students. This is in line with the finding by Myles et al. (1998) that learners use formulaic language, and some are better at memorising it. In addition, the visual learning of formulaic language was dominant in both groups of students. However, students still lack instruction regarding distinguishing between formulaic language use in the spoken and written registers. This leads to textual analysis results which show that the student written texts in this research include more spoken language features, as more verbalised formulaic language expressions are used to refer to the real world (see section 7.1).

Additionally, language learning in the research context is more driven by the testing culture based on Chinese-designed examinations, than by the native norms of English. In other words, the students are generally more interested in getting higher scores in the exams than chasing after or learning to use any native norms of English.

7.3.2 Using formulaic language

With regard to using formulaic language, three themes emerged from the perceptual data, that is, the process of using formulaic language, formulaic language use strategies and formulaic language use in English writing. Based on the analysis of the interviews,

formulaic language use strategies are intertwined with the process of formulaic language use, and there is not a clear-cut boundary between the process and strategies employed in using formulaic language. The process often consists of continuous evolving stages, whereas the strategies can be more specific and used to solve certain issues.

According to the statements that interviewees provided, both groups shared a similar process in using formulaic language, which involved the translation of English expressions into a Chinese equivalent, and then Chinese to English again (see Figure 6.22). Thus, the process of using starts with the comprehension of the input materials, at this stage, the interviewees need to translate the language of the input materials (English) into their L1 (Chinese) to comprehend them, and this is followed by the process of retrieving formulaic language from memory, usually the interviewees find equivalent English expressions which match the comprehended Chinese expressions. Sometimes, the interviewees find more alternatives for the equivalent English expressions.

With regard to formulaic language use strategies (see section 3.1.3), according to the interviews in the study, cover and communications, as language use strategies, were not reported by the students at all, whilst retrieval and rehearsal seemed to be the most popular language strategies used by the students. Language transfer and rhetoric techniques were usually used in the process of retrieving formulaic language. Language transfer, for example, was used as an analytic tool in using formulaic language: once they had the Chinese expression, they then would seek for the equivalent in English from the memory (see section 6.2.3). Also, they used contrastive rhetoric as a strategy in using formulaic language, for instance, influences from their Chinese writing were brought to their English writing, for example, they claimed that they consciously avoided common formulaic language usage in order to be creative in English writing (see section 6.2.3).

The other popular formulaic language use strategy is rehearsal. Students stated that efforts were made to prepare formulaic language in mind which could probably be used under certain circumstances, including formulaic language used in their English writing.

In other words, students had an awareness of the use of formulaic language across various registers and genres.

In addition, students felt that formulaic language was used differently in spoken and written language, as well as writing for different genres, for instance, argumentative writing and narrative writing. However, it seems that, for both groups, a larger number of participants noticed more generic differences than register differences, for example, the differences between oral and written language (see 6.2.3.4).

This finding is in accordance with the results from the textual analysis, which shows that for both groups of students written texts featured the spoken registers heavily. The distinction between the spoken and written language is usually made by the identification of 'speech-like' items or as language within an 'oral-tone' and has been defined in the literature in comparison with a native norm or according to the researchers' intuitions (Cobb, 2003; Granger, 1998; Lee and Chen, 2009; Leedham, 2011). Previous studies on Chinese students' writing also found that Chinese writers tend to use forms that are more prevalent in speech compared to their non-Chinese counterparts, (Mayor, 2006; Leedham, 2011).

Leedham (2011) gives two possible explanations for this phenomenon: First, the Chinese writers' greater use of informal lexical items and chunks is due to their greater familiarity with spoken than written English. Second, it is linked to the genres of the texts examined, which mostly are argumentative essays requiring no research or preparatory reading. As students were asked to give their opinion on topics of general knowledge, the need to engage in formal and/or academic language was limited.

The findings from this study show the intricate relationship between the written and spoken formulaic language used by the students in the research context.

On the one hand, the students had to pay more attention to the written form of formulaic language, as they had to achieve correctness and accuracy in exams, including

filling the gaps and short English essays, where they were asked to keep the formulaic language in place.

On the other hand, they learned formulaic language mostly outside the classroom and mainly through visual learning methods and did not receive enough instruction from teachers or have a sufficient awareness of spoken and written language differences.

This leads to confusion in terms of formulaic language use in different registers, especially in the oral and written forms. In fact, they used some oral and informal forms in their writing, for example, they used conversational expressions, like *nice to meet you* in their texts; they also used formulae (*no pain, no gain*) and contractions (*I think it's; there's no doubt*), which are usually used in the spoken language. This is also an additional explanation to the phenomenon that Chinese EFL writers' heavy reliance on spoken language in their writing despite the lack of speaking practice.

As for formulaic language use in English language writing, students reported that they chose and used formulaic language according to the genres they produced, for example, they only used the writing templates in English writing for assessments, not for personal purposes. Students also mentioned the use of famous quotes, idioms and fixed expressions and large chunks of language from the writing of others (see section 6.2.3).

Students in the survey were also asked about their viewpoints on the formulaic language used by themselves and their peers. For their own use of formulaic language, it was reported by the students that they tended to use formulaic language with slight changes in order to avoid using memorised formulaic language, as the repeated use of formulaic language in one text was regarded as an indication of poor quality writing, which was not acceptable to students from either group. However, in terms of the repeated use of formulaic language across the students' written texts, students from the two groups had ambivalent attitudes, as they had to follow the schematic structure for the writing task. In other words, they were not sure about whether the influence was positive or negative.

To summarise, for the students in this study, the process of using formulaic language starts with the comprehension of the input materials, at this stage, the interviewees need to make connections between the language in the input materials (English) and their L1 (Chinese). This is followed by the process of retrieving formulaic language from the memory, and usually the interviewees find equivalent English expressions which match the comprehended Chinese expressions. Sometimes, the interviewees found more than one alternative English expression. Regarding formulaic language use strategies, the most popular strategies reported by the students were retrieval and rehearsal. Language and contrastive rhetoric techniques were used as analytic techniques to retrieve formulaic language at the time of use. In terms of rehearsal strategy use, students did prepare formulaic language for use in advance. Students showed a greater awareness of genre than register (spoken and written language) differences. In English writing, the repeated use of formulaic language, on the one hand, if it is in one text, is regarded as a bad thing and is not considered to be acceptable by students; on the other hand, if it is repeatedly used in different texts in the class, the students remain unsure whether this is positive or negative.

7.3.3 Students' comments on teaching

The data gathered from the interviews concerning the students' comments on English language teaching can be roughly divided into three sub-themes: textbooks, teaching methods and teachers.

As for the students' comments on their textbooks, it is worth noting that the authority of the textbooks become less important for the Year 3 students. Most interviewees from Year 1 gave positive comments on the content of textbooks with regard to meeting their formulaic language learning needs. They used words such as "helpful, satisfied and excellent". Only one interviewee was not satisfied with the textbooks, because of the rather limited content. The Year 3 students' attitudes towards English textbooks, however, seemed to be more varied and critical, as some of them reported that the content with regard to formulaic language in the textbooks was "not much or enough".

Comparing the comments from both groups, the Year 3 students presented more complex attitudes towards their textbooks and were keen to explore the English language through other means than just from the textbooks authorised by the university.

With respect to the teaching of formulaic language, in general, the students reported that they were satisfied with the content taught in the classroom. However, according to the students' comments on teaching methods, especially on formulaic language teaching (see sections 6.2.3), it seems that the grammar-translation method and audio-lingual method have been the dominant methods adopted in the teaching at the research site, especially in the teaching of the Year 1 students. For the Year 3 students, the teaching seemed to pay more attention to general language use and meaning-making.

The following issues could be related to the teaching methods on the research site: first, the direct results of these methods of English teaching might well lead students to only focus on the form of language at the expense of other aspects of English language. The focus of the teaching of formulaic language was put on the form of language rather the function, or integration of form and function. For instance, the teachers stressed the importance of formulaic language in class, but did not make it clear whether formulaic language was important for language learning or for students to pass exams. This unspecific instruction left the students with an ambivalent explanation of the functions of formulaic language. Second, there are not enough communicative events between teachers and students in the class, as formulaic language teaching stressed accurate memorisation as well as some memorisation techniques. Teaching seems to focus only on certain types of formulaic language: writing templates and phrases, and especially concentrates on the explanation of the grammar in these items. Third, there is lack of access to authentic English material and this has been neglected in the teaching, since some students realised the gap between the formulaic language learned in the class and the "authentic" language used in "real" life across different genres (see section 6.2.3).

When it comes to teachers, the students talked about domestic and foreign teachers separately. The instruction of formulaic language was explicitly given by domestic teachers across all levels, from the middle and high school to university, and explanation was used as the main approach. However, based on the interviews (see section 6.2.3), we can conclude that the emphasis on memorising formulaic language in English teaching at university is not as strong as it is in middle or high schools. Furthermore, there is no consensus or communication between students and teachers on the issue of "what types of formulaic language should be taught to the students" (see Example 18.2). In addition, domestic teachers were oriented by native norms in their teaching of formulaic language and the interviewees reported that the native norms were usually simplified.

On the other hand, foreign teachers give instructions on formulaic language in both explicit and implicit ways. They not only teach and encourage students to use idiomatic expressions explicitly, but also make communicative exchanges happen and teach formulaic language in an implicit way when they talk to students in English in real communication. Moreover, foreign teachers are recognised by the students as native speakers using "authentic" English. With this strong native-norm drive, the teaching was exclusively aimed at the native-norm, which is mostly up to the individual foreign teacher's perception and proficiency. The foreign teachers, for example, often criticised the content in the textbooks as the language is not used in real communication. Additionally, compared to domestic teachers, foreign teachers tended to be more confident in English formulaic language use.

Based on the discussion in this section, we can summarise that the students offered well-rounded comments on the teaching of formulaic language that they had received, specifically in terms of the teaching materials (compulsory textbooks), teaching methods and approaches as well as the teachers' personalities. For the students, with the development of language proficiency and communicative needs, compulsory textbooks did not satisfy their needs any more, as they were keen to look for authentic English

materials outside the classroom which could be used in real communication. As for formulaic language teaching, we can note that most teaching at this site is still strongly grammar-and form-focused. The dominant teaching methods are grammar-translation and the audio-lingual method. This had the effect that formulaic language teaching at the research site was driven towards improving correctness and accuracy in formulaic language use. On the other hand, communicative exchange is scarce between teachers and students in terms of formulaic language use and formulaic language teaching. In addition, the native-norms of English played a role in English language teaching, since the ultimate goal for both students and teachers was to encounter authentic English. It is worth noting that terms, like “native”, “native speakers”, “authentic English”, “correctness”, “appropriateness” and “accuracy of English use” were not clearly distinguished, especially by the students and domestic teachers. Foreign teachers at the research site showed strong native-norms in their teaching in contrast to the domestic teachers, who often used simplified and varied expressions.

7.4 Summary

This chapter provides a more comprehensive discussion of the three sets of research questions. Based on the analysis and discussion of textual data and perceptual data, the student conceptualisation, learning, use of formulaic language and student reflections on formulaic language teaching have been depicted.

The most significant finding is the disparities that lie in the distribution of formulaic language identified by two approaches: the corpus linguistic approach and the students’ perceptions. The disparities of the results reflect that there is lack of common ground between the researchers’ and language users’ (or students’) views on formulaic language.

As for the student conceptualisation of formulaic language, first, although the students reported some views, and indeed, the Year 3 students showed a deeper and broader understanding of formulaic language. Almost all the students’ comments were limited to a relatively shallow level. Without any concrete examples to support their ideas or

opinions on their conceptualisation of formulaic language, the students just talked about some ambivalent features, or simply repeated their teachers' instructions about formulaic language in the classroom.

Second, the students' perceptions of formulaic language did not always match with the researchers'. In the study, for instance, some students mentioned that the use of formulaic language under certain circumstances can produce humorous effects, which is a theme which has long been absent in formulaic language and English language learning and teaching research. Students paid more attention to the fixedness of formulaic language than the frequency factor; this is probably caused by the different views on language between students and researchers: researchers may have an overview of the rules underlying the language, whereas the students only care about formulaic language use in single situations.

With regard to the learning of formulaic language, the general learning process for formulaic language was shared by the students from both groups. This starts with attention or noticing, and then memorisation, followed by a review or retainment of formulaic language, and ends up with the retrieval or recall of the formulaic language at the time of use. The main learning strategies reported by the students are memorisation, accumulating the items and communication. The Year 3 students showed that they used more extensive strategies, such as comprehension, classification and independent learning, which may indicate that more implicit learning of formulaic language happened among these students. In addition, the visual learning of formulaic language was dominant in both groups of students.

In terms of using formulaic language, the process of using formulaic language starts with the comprehension of the input materials. At this stage, the students need to make connections between the language in the input materials (English) and their L1 (Chinese), and this is followed by the process of retrieving formulaic language from the memory, usually the students find equivalent English expressions which match the comprehended Chinese expressions. Sometimes, the students find more than one

alternative English expression. Regarding formulaic language use strategies, the most popular strategies reported by the students were retrieval and rehearsal. Language and contrastive rhetoric techniques were used as analytic techniques to retrieve formulaic language at the time of use. For the rehearsal strategy use, students did prepare formulaic language for use in advance. Students showed their awareness of genre and register differences. In English writing, the repeated use of formulaic language, on the one hand, if it is in one text, is regarded as a bad thing by the students and is not considered to be acceptable; on the other hand, if it is repeatedly used in different texts in the class, the students remain unsure whether this is positive or negative. This shows that students had an awareness of the use of formulaic language, specifically in English writing, however, their statements in the survey reflect their lack of insight into the issue of formulaic language use.

As for formulaic language teaching, we can note that most teaching at this site is still strongly grammar- and form-focused with a strong rote-learning influence. The dominant teaching methods are grammar-translation and the audio-lingual method. This had the effect that formulaic language teaching at the research site was driven towards improving correctness and accuracy in formulaic language use. On the other hand, communicative exchange is scarce between teachers and students in terms of formulaic language use and content of formulaic language teaching. Additionally, the language learning in the research context is more driven by the testing culture based on Chinese-designed examinations than by the native norms of English, although the native-norms of English played a role in English language teaching, as the ultimate goal for both students and teachers was to encounter authentic English. However, the students are generally more interested in getting higher scores in the exams than in chasing after or learning to use any native norms of English. It is worth noting that terms like “native”, “native speakers”, “authentic English”, “correctness”, “appropriateness and accuracy of English use” were not clearly distinguished by the students and domestic teachers. Especially in terms of formulaic language with Chinese English and China English, there is no explanation or caution from either home teachers or students. Foreign teachers at

the research site showed strong orientations of native-norms in their teaching in contrast to domestic teachers, who often only encouraged students to used simplified and varied expressions.

In conclusion, while there are some limitations to this research, it is hoped that this explorative description can shed light on the reconceptualisation of formulaic language, the learning and use of formulaic language in the EFL context. In turn, it is also hoped that this research will lead to some implications for the teaching of formulaic language, especially in the EFL context.

Chapter 8 Conclusion and implications for teaching

This final chapter will present a summary and clarify the contribution of the current research on formulaic language by reviewing the research rationale, the literature review and also the formulation of the research questions. This will be followed by revisiting the findings regarding the research questions and focus. I then restate the limitations of the research and make some recommendations for future work. Before arriving at the final remarks, the implications for language teaching will be discussed based on the findings of this research.

8.1 Research rationale

This study aims to provide an accurate description of formulaic language in Chinese university students' written texts, as well as attempting to explore the learning, use and teaching of formulaic language at the research site. The study builds upon an innovative design and provides new insight into formulaic language use in EFL settings to make its own theoretical and methodological contribution to the study of formulaic language.

The theoretical position of formulaic language is first discussed in Chapter 2. Previous research on formulaic language mostly focused either on describing its linguistic features (by the speaker-external approach), or on the processing of formulaic language (by the speaker-internal approach) (cf. Wray, 2002; 2008; Schmitt, 2000; Biber, Conrad and Cortes, 2004; Hyland, 2008 among others). Therefore, numerous terms were used in different research strands to mean "formulaic", but they all showed different biases. For example, studies on multi-word units, N-gram, clusters and lexical bundles are mostly based on a corpus linguistics approach, which relies on frequency criteria to extract word combinations from a corpus, usually from a large corpus comprised of native or non-native speakers of English; while, from the perspectives of psycholinguistics, research seems to focus on the processing of formulaic language, or the perceptually salient chunks in the language production (see section 2.1).

Second, some researchers argue that language description and language explanation should be taken into consideration at the same time (Bhatia, 1993, 2004; J. Flowerdew, 2002b; Hüttner, 2005; 2007; 2008). Previous research does, in fact, not appear to reflect this point. Instead, more research has been done on the contrastive analysis of learners' language and native-norms, while there is little research linking the description of formulaic language used by a small community in the EFL setting to students' perceptions (see section 3.4).

Drawing on both points, the prerequisites of formulaic language for this research can be met. These are: 1) Multiword psycholinguistic units, which can be stored holistically or with processing advantage, i.e. it is learned, memorised as a whole for the learner, and/or retrieved with greater efficiency than other linguistic strings at the time of use, regardless of native language norms during identification; 2) Linguistic cluster, which meets a series of selective criteria in a corpus.

Methodologically, compared with spoken language, the written language has been considered secondary in traditional second language learning and formulaic language learning research. Although there are some prior studies on formulaic language in written corpora, either the corpora are large and comprised of texts written by different writers, or the data for the corpora were collected at one point in time, rather than were compiled by the student written texts consisting of the identical writers or a truly "longitudinal study" (Paquot and Granger, 2012).

On the basis of these considerations, I therefore draw on genre studies, which focus on the study of conventionalised language use in a community (J.Flowerdew, 2002b; Hyland, 2004; 2009; Swales, 1990), and then link it to formulaic language. Both genre studies and formulaic language studies examine conventionalised language at macro- and micro-level respectively (Hüttner, 2007). Hereby, I include writing templates, or schematic structures in this research as formulaic language at macro-level. Next, I am going to review the research questions, methodology and findings in turn.

8.2 Research questions and methodology

The present study is an explorative investigation of the learning and use of formulaic language by Chinese university students in their English written texts. With mixed-methods including corpus linguistics and interview data, this study aimed at providing a precise description of the formulaic language used in their English written texts. With regard to the analysis of linguistic features and discourse functions of formulaic language, formulaic strings were respectively identified by the students and the lexical bundles extracted from the learners' written corpora. Three sets of research questions were established to achieve the purpose of the research as follows:

(1) To what extent do Chinese university students use formulaic language in their written English?

a) What are the main structures of formulaic language used by these learners?

b) What are the main discourse functions of formulaic language used by these learners?

c) What is the relationship between the distribution of structural and functional categories of formulaic language in the learners' written texts?

d) How is formulaic language used differently in the written texts of Year1 and Year 3 university students?

(2) What do Chinese university students perceive formulaic language to be?

a) To what extent is this perception different in Year 1 and Year 3 students?

(3) How do Chinese university students perceive the learning and use of formulaic language?

In Chapters 6 and 7, two types of data collected from each research method were introduced: textual data used to assemble the student written corpora and perceptual data gathered through two rounds of semi-structured interviews; and then findings in relation to each set of research questions were analysed and presented.

For the textual data, the written texts were collected during one semester (16 weeks, 4 tasks) from the Year 1 group and the Year 3 group respectively. The Year 1 corpus used in this research contains 56,058 words and 300 texts, whereas the Year 3 corpus consists of 64,186 words and 251 texts. The identification of formulaic language is based on two approaches: first, formulaic strings were identified by the students, which is based on the speaker-internal approach. The informant writers were asked to mark the language as it is learned, memorised as a whole, or as a holistic unit, rather than constructed word by word at the time of use. The second approach was to examine the four-word lexical bundles automatically extracted from the two corpora. The threshold frequency was normalised and set at 2 and 3 occurrences in the Year 1 and Year 3 corpus respectively. The bundles had to appear at least 5 texts to avoid "individual quirks". The same or overlapping formulaic language (i.e. formulaic strings and four-word clusters) were identified by both approaches and it was termed as "core formulaic language" in this study. Also, the terms "shared formulaic strings" refer to the same or overlapping strings identified in both corpora; and "shared four-word clusters" refer to the same overlapping four-word clusters identified in both corpora (see sections 5.8.1, 6.1.3 and 6.1.4).

As for the perceptual data, it was collected from two rounds of semi-structured interviews with the same interviewees. The first round was conducted at the beginning of the semester and the second round was taken out at the end of the semester. 6 student writers from the Year 1 and 6 student writers Year 3 groups volunteered to take part in the rounds of interviews (12 interviewees in total). Based on the integration of bottom-up and top-down analytical approaches, interviews were analysed under four themes in relation to the research questions. These are: the students' understanding of formulaic language, formulaic language learning, using formulaic language, and formulaic language teaching and issues related to the contextual factors.

8.3 Findings

By integrating two types of data, it is possible to make a justified description of formulaic language learning and use in the research context.

8.3.1 Reconceptualisation of formulaic language

The reconceptualisation of formulaic language is realised through the findings of two types of data analysis: the textual data analysis based on two student written corpora; and the perceptual data analysis from two rounds of semi-structured interviews with students.

In the textual analysis, I found both types of formulaic language, i.e. formulaic strings identified by the students and four-word clusters extracted based on the corpus linguistics approach share some general trends in terms of structure and function. In order to offer a description of identified formulaic strings and four-word clusters, the researcher then assigned those strings and clusters to the linguistic structural and discourse functions. For the structural level, all the strings and clusters fit into three broad structural categories: phrasal structures, clausal structures, and other expressions. Phrasal structures are the most identified, followed by clausal and other structures respectively. Furthermore, the VP-based (mainly actively phrasal verbs) are the most identified phrasal structure, whereas under the clausal structure category, the WH-clause and the to-clause are the most identified structures. As for the functional categories, all identified strings and clusters share three broad functions, that is, ideational, textual and interpersonal functions. The majority of formulaic strings and four-word clusters perform an ideational function (description and action), usually with VP-, NP- and PP-based phrasal structures. This is followed by the textual function, most commonly with PP-based or other structures. The interpersonal function is the least frequently identified, mostly with clausal structures.

However, there are some differences between the findings regarding these two types of formulaic language, which reflects the divergence between the speaker-internal

approach (psycholinguistic approach) and the speaker-external approach (corpus based approach), in terms of the theoretical background, methodological considerations and research focus. For example, there are no clusters with ill-formed or template structures found in either corpus, which again, on the one hand, shows that individual idiosyncratic items or idioms like formulae are too infrequent to be investigated by the corpus linguistics approach due to the slight changes in each occurrence. On the other hand, it also highlights that the corpus linguistics approach adopted a clear-cut method in examining formulaic language, with regard to the continuity, frequency, length and text dispersion of the combination. This can turn into a limitation of the research.

With the perceptual data analysis, students listed their learning preferences for different types of formulaic language in the interviews, with the top three types being phrases, idioms and fixed expressions, templates and/or phatic expressions. This may indicate that phrases, idioms and fixed expressions and writing templates are widely used and favoured by the students when they need to construct written texts, which reflects the role of formulaic language as building blocks for writers, especially for student writers (Hüttner, 2007; Hyland, 2008).

Evidence from the interviews shows that the students focused primary on the form of formulaic language, as they referred to fixedness as the most distinctive feature of formulaic language. Other features like frequency, with a literal or figurative meaning were mentioned occasionally. In terms of content, students talked about the specific types of formulaic language, for example, phrases, idioms and fixed expressions, but were not able to give detailed descriptions or examples. This could indicate that in their views, types of formulaic language with features like a relatively fixed form with literal meaning, such as formulas or proverbs, are easy to learn, but in contrast, items that are flexible or allow variations with figurative meaning, such as idioms and phrases are difficult for them.

From both the analysis of textual data and perceptual data, it can be seen that the formulaic language that used, recognised and mentioned by the students in these two

types of data covers both micro-and macro-levels. In other words, formulaic language can range from the level of words, for example, derivations or inflections of a word, conjunction words, or hyphenated compounds, to phrases and sentences (for instance, formulae, clauses), up to the macro-level, for example, the schematic structure for a text (see sections 6.1.2 and 7.1.1).

I developed the term "core formulaic language" alongside the textual data analysis. It refers to the same or overlapping or formulaic strings and four-word clusters pairs in one corpus (see section 6.1.5). The total number of core formulaic language pairs in the Year 1 and Year 3 corpora is the same; there are 15 pairs in each corpus, even though these writing tasks are on different topics. Furthermore, the analysis shows that the structure boundaries of formulaic language can be blurred or fluid, as for both formulaic strings and four-word clusters can either extend to the other structure elements or not, however, the function of the formulaic language remains the same (see section 7.1.3).

In conclusion, there are still discrepancies in the recognition of formulaic language, between the students' perceptions and frequency-based corpus linguistics approach in terms of structural and functional categories. The reason behind this could be the different stances of the researcher who relies on the corpus-based approach, and the users of the language, as they do not share a focus on language production. For example, the users in this research are students and besides expressing their own meaning through the written texts, they had to pay a certain amount of attention to ways of structuring their texts, whilst the corpus linguistics researcher only saw the final product of the writing process, without thinking about the structure of the texts by putting him/herself in users' (in this case, students') shoes.

Overall, most participants acknowledged the existence of formulaic language in their English, and recognise that it is relatively popular (Erman and Warren, 2000; Foster, 2001; Nattinger and DeCarrico, 1992). However, it is surprising that students are more impressed by the holistic features or fixedness of formulaic language, than the frequency factor (see Chapter 2). They also considered formulaic language as playing a

versatile role in their English use, helping to improve correctness and appropriateness in their language production. Furthermore, students paid attention to the role of formulaic language in facilitating linguistic construction and achieving communicative purposes. In addition, it is noted that the students paid attention to the possible literary effects produced by using formulaic language, for instance, humour, which have been often ignored widely by researchers of EFL teaching and learning.

8.3.2 Formulaic language learning

In the student self-reports on their formulaic language learning, several issues were reflected.

Firstly, as mentioned above, the majority of interviewees realised the existence of formulaic language, however, not all the students in this research thought that formulaic language was important for their English learning. In other words, not all students in the study have the awareness of formulaic language in English.

Secondly, language learning in the research context is more driven by Chinese-designed English tests rather than by the native norms of English, in other words, generally, the students are interested in getting higher scores in exams rather than chasing after any native norms of English. Thus, the students have been trained to complete writing tasks under exam conditions according to the assessment rubrics and sometimes, the students try to write to impress the graters of the exams rather than write to express their viewpoint. A tentative explanation for this could be that in order to formulate ideas as well as achieve appropriate, accurate and correct use of language, the students have been taught to learn, memorise and use building blocks in the English written texts (Hüttner, 2008), as formulaic language can ensure correct, grammatical and appropriate use of English.

Third, interviewees shared the general process of formulaic language learning, which starts with attention or noticing, and then memorisation, followed by the review or retainment of formulaic language, and ends up with the retrieval or recall of formulaic

language at the time of use. It is notable that memorisation still plays an important role in formulaic language learning among the students in this research. Nevertheless, it seems that most interviewees reported that they just left the learned formulaic language alone or unanalysed. The question remains whether the learned formulaic language is later segmented or analysed (Peters, 1983).

As for the learning materials, in the research context, students tend to rely heavily on textbooks and teachers' instructions in the classroom, and visual learning materials outside of the classroom. Therefore, explicit learning plays an important part in formulaic language learning on the research site, with focus on the form of language.

Finally, not all interviewees were able to report specific strategies for each kind of formulaic language learning, which could indicate that they were not aware of some strategies used in learning. This may also be caused by a lack of explicit teaching or instruction from their teachers.

8.3.3 Using formulaic language

Based on the findings from the interview data, the process of using formulaic language starts with the comprehension of the input materials. At this stage, the interviewees have to transfer the language of the input materials (English) to their L1 (Chinese). They then retrieve formulaic language from the memory and usually as the interviewees reported that they find equivalent English expressions which match the comprehended Chinese expressions (see section 6.2.3). Sometimes, the interviewees find more alternative English expressions. During this process, transfer between two languages and contrastive rhetoric techniques were used as analytic tools. This finding resonates with the argument by N. Ellis (2012), that adult L2 learners already know about the existence of language units, categories, and linguistic structure. Thus, the second language learners expect that there will be words and constructions in the L2 that correspond to such word classes and frames. Ellis (2012: 40) also states that once the learners have identified, or even once they have searched out and actively learned such key

vocabulary, they “are more likely therefore to attempt creative construction, swapping these elements into corresponding slots in frames”. However, the role of L1 has been neglected for a long time by research on formulaic language in second language learning, especially in terms of adult learners.

As for the repeated use of formulaic language, except for templates, or schematic structures, if it is in one single text, it is regarded as an indication of unsuccessful language use and is not considered to be acceptable by the students. Regarding the repeated use of formulaic language in different texts in the class, the students stated that they have to follow “the abstract structure,” in each section of the text, as these schematic structures are compulsory in their English writing requirement and rubrics for the assessment. Furthermore, the students supported the claim that they cannot control others’ language use and they seemed not be bothered with the issue. This is probably led by the still popular rote-learning in the research context, which mainly adopts audiolingualism in the classroom, with its stress on repetition and structural drilling. It is clear that formulaic language has maintained a place in contemporary classroom practice in Chinese universities, however, often further concrete and specific instructions are not given to the students.

8.3.4 Student comments on formulaic language teaching and contextual factors

The students offered well-rounded comments on the teaching of formulaic language that they had received, specifically in terms of the teaching materials (compulsory textbooks), teaching methods and content as well as the home and foreign teachers. For the students, with the development of the language proficiency and communicative needs, the only compulsory textbook is no longer satisfied their needs in English language learning and using; and the students crave authentic English materials that they can use in real communication. As for the teaching methods and content, it is noted that most teaching is still strongly grammar-and form-focused. In both teachers’ and students’ perceptions, formulaic language considered to help improve fluency and correctness in the students’ English rather than make their English more nativelike.

There are still some gaps between learning needs and teaching, for example, a consensus of opinion on the question “what types of formulaic language should be taught in the classroom”, was not reached by the students and teachers (see Chapter 4).

8.3.5 Group differences

With regard to the group differences in formulaic language use, it involves the analysis of two types of data: textual data and perceptual data.

The textual data reflects the formulaic language used by the students from two groups. The initial results from two identification approaches show that the senior students tend to use more and longer formulaic expressions. However, as the reason could be the Year 3 students produced longer texts than the Year 1 students. After two calculations to normalise the quantity, however the mix results show that it is not necessary that more proficient users would use more formulaic language. In terms of structure, overall, the Year 3 students used more formulaic language than the Year 1 students in most structural categories. Except for the clausal structural category in the student identified formulaic strings. For the functional aspect, although the results from both approaches are divergent, the findings also show some convergent aspects, for example, the most pervasive form of formulaic language has an ideational function; the Year 3 students generally used more formulaic language than the Year 1 students across all three main discourse functions: ideational, textual and interpersonal. These findings seemingly contradict Hyland’s (2008) study, in which he indicates that there is relationship between the use of lexical bundles and writing expertise and that more advanced writers tend to rely less on lexical bundles.

Also, the researcher used term “shared formulaic strings” and “shared four-word clusters” to refer the same or overlapping strings and four-word clusters identified in both corpora respectively. There are 21 pairs of formulaic strings and 20 pair of four-word clusters identified in the two corpora in this study. This could provide the evidence

that captures the formulaicity, to some extent, in the students' written texts (see Chapters 6 and 7).

In terms of the perceptual data, the interviewees from Year 3 showed a broader understanding of formulaic language and expanded the vision of formulaic language, by providing examples of proverbs (example 4), as well as mentioning slang and phatic expressions, such as greetings as possible forms of formulaic language (example 5); they also started to think about some formulaic language with opaque meanings (example 6), which could be an obstacle to comprehension (see Chapter 2).

There is not a noticeable difference between the Year 1 and the Year 3 students' perceptions of the roles of formulaic language. For both year groups, formulaic language undertakes both negative and positive roles in their English use, but in general, the positive roles of formulaic language outweigh the downsides of formulaic language use. However, the Year 3 students tended to see the role of formulaic language as a dynamic one, as one interviewee pointed out that formulaic language was very helpful in the early phases of her English learning, yet, later on, she found it limited her writing to some extent when her English reached a certain level.

8.4 Limitations and recommendations for future research

Despite the fact that this research has employed a mixed methods design, which aims to reduce the weaknesses of each research method, some limitations remain.

First, regarding the qualitative strand of this study, the small number of participants and the single setting of this study as a case to some extent restricts its transferability to other contexts, although the researcher provided reasons why she chose this university for conducting fieldwork (see sections 4.3.1 and 5.3). The participants all share very similar linguistic, social, and cultural background, as they are all in the same age group, and from the local area, where the dialects are highly similar. It has to be acknowledged that if a different university were chosen, for example, a top university in China, this

could result in different findings, as the students may come from a wider range of background in language experience and social cultures (see section 5.3).

Second, as for the quantitative strand, student written texts were collected over one semester (16 weeks) and consisted of two corpora. By using a corpus linguistics approach, several issues involved in the corpora compiling and operationalization of the analysis emerge (see section 6.1.2). For example, the results may not be compared directly across studies, because of the differences in the size, the threshold frequency, distributions, length of bundles, and the topics of the written texts (see sections 3.2.2 and 6.4). In this study, the sizes of both corpora are relatively small, and therefore the results may not be compared directly with those from larger size corpora (Hyland, 2008; Biber, 2006; Chen and Baker, 2010). However, through the socio-cultural and psycholinguistic description provided in this study, it is possible to provide more insightful information to understand research on formulaic language in an EFL context.

The mixed methods design itself also presents some weaknesses (Creswell, 2014; Bryman, 2012). One such limitation is the integration of the results from two strands, which probably leads to difficulties with comparing results from two analyses using data from different forms. Although the research makes the importance of the data and the interaction level of data obtained by both methods clear (see section 5.2), some questions still need to be addressed thoroughly. For example, the student identified formulaic strings and lexical bundles extracted from the corpora show disparities and the researchers only can give a tentative explanation on the basis of previous studies, rather than investigating the issues directly with the participants. In addition, as the mixed methods design is employed in this study, the whole process, including the research design, data collection and analysis, is relatively time-consuming.

Another limitation is that due to time restrictions, the teachers' perceptions of student formulaic language use and learning could not be addressed in any detail. However, an inclusion of these views would add to the conceptualisation of formulaic language by researchers, teachers and students. In this way, the understanding of formulaic language

can be coordinated in teaching, learning and research, and the results from these proliferate formulaic language research can be used and can guide teaching in various contexts. In other words, formulaic language research should further take the perceptions of students and teachers into account.

8.5 Contributions and implications

By endeavouring to answer the three sets of research questions, this study has not only contributed to a better understanding of how formulaic language is used and perceived by Chinese university students, but also reflected certain issues of English formulaic language learning, using and teaching at the research site. Since it is an ordinary medium-ranked university and there are numbers of universities in China in more or less similar situation, its issues are representative.

First, according to the findings of the textual analysis in this study, the broad structural and functional categories, whilst sharing some distributional trends, are noticeably diverse in terms of the distribution of the structural and functional sub-categories (see Chapters 6 and 7). On the one hand, this is caused by different methods based on different theoretical roots. The formulaic strings are psycholinguistically salient to the language user with in a community, while the clusters are based on a series of criteria outside of the language users which relates to occurrence and a limited length of word combinations. For example, there are some clusters identified with a clausal structure, which are usually used to help construct a sentence. These are unlike some strings, which could function at a micro-level to construct a sentence, or a macro-level as a schematic structure for a whole text. On the one hand, formulaic strings are often associated with individual idiosyncrasies, such as changeable forms and ungrammatical deviation, which contribute to the low frequency and are rejected by the corpus linguistics (see section 6.1 for discussion on formulae). These can all be factors that resulted in the discrepancies in the findings between the two identification approaches. The mismatch reflects the mismatch in the perception of formulaic language by researchers and students. For students, the formation of formulaic language seems to

be more complicated and influenced by many factors, such as language transfer, contrastive rhetoric, cultural influences, as well as the purposes of writing (i.e. write to express or write to impress). Whereas for the corpus researchers, most times only examine the final language manifestations regardless of the factors that may influence formulaic language use. In the future, more research will need to be done in the field of formulaic language in EFL settings to triangulate the conceptualisation of formulaic language by researchers, teachers and learners.

Second, with regard to formulaic language learning, according to the interviews, students reported that fixedness is a distinctive feature of formulaic language, but not factors like frequency or opaque meaning. They claimed that relatively fixed formulaic language, with a literal meaning, is easy to learn, where those items with flexibility or variations, and with figurative meanings are more difficult to master. More explicit instruction should be given on those difficult items rather than on these easy items for students. On the other hand, from the analysis of the formulaic strings and clusters, I could establish that both strings and clusters can have incomplete structures. However, the formulaic strings with incomplete structures are either semantic units or they function a grammar reminder. They both make sense and are meaningful to the language users, while the clusters are not necessary in this case.

In addition, the issues regarding the formulaic language learned both in the classroom and outside class emerged in the student comments on formulaic language teaching in the research context. As for the formulaic language teaching in the classroom, it is significantly influenced by rote-learning and a prescriptive approach. With the strong rote-learning orientation, audiolingualism and grammar-translation methods were largely used in the classroom, which often stress repetition and structural drilling. This led to the result that the general patterns of formulaic language use do not significantly change or progress from the junior to the senior year student written texts. In the classroom, the overly prescriptive approach to formulaic language teaching was adopted and it can easily imply to students that they only have to learn basic textual structures to

create a genre that meets the expectations of a particular discourse community (Dudley-Evans and St John, 1998). Teachers should be cautioned against turning genre-based instructions into a "formulaic type of instruction" in which students are simply instructed to manipulate certain features and writing to impress the teachers rather than to express their meanings (Reppen, 1995).

The rote-learning and prescriptive approach used in the formulaic language in the classroom, also influenced learning outside the classroom. There are two problems: first, the students lack guidance on selecting materials for English language learning outside the classroom. They mainly relied on audio-visual sources, often without critical assessment of these sources (see section 6.2). This probably leads to the second problem that the students in this research generally lack an awareness of the differences between oral and written language. This is also reflected in the findings of the textual analysis, which shows that formulaic language used in the students' written texts from both year groups featured oral language.

Further research and teaching on formulaic language that has been suggested by this study also includes some interesting cases regarding China English and Chinese English (see section 7.1 and 7.4). Examples of China English are *tomb sweeping day*, *Qing Ming Festival*, *the dragon boat festival*, *Pure Brightness Festival*, and *harmonious society*, and they all contain Chinese cultures and reflect features of the social environment with Standard English grammar; whereas Chinese English expressions like *waste-time and waste-energy thing*, are actually word-for-word translations of Chinese expressions "*wang fei gong fu* (枉费工夫)" and are ungrammatical in English. However, instructions regarding these two are not sufficient. The teachers gave neither an explanation of the differences between these two types of expressions, nor did they try to position China English in the English language classroom and so support their students in being able to communicate appropriately. Especially, the students had very limited practices with native speakers of English at the research site. More research is needed to examine the

position of this type of English formulaic language in English language learning and teaching, especially in the EFL classroom.

Furthermore, this study also represents some common problems with university English writing in China in general. My data suggest that the written texts in this research are nowhere near the level of so-called “academic writing”, as both written corpora contain oral language features, subjectivity and lack textual plurality (Moore and Morton, 2005). The reasons might be speculated on as follows: first, the nature of the writing tasks in this research is different from traditional academic writing. These tasks are mostly based on recounting personal experience or on general knowledge of the real world and do not ask the students to cite or refer to work from reading. Second, as Leedham (2011) notes, Chinese students are primarily exposed to sentence-level analysis, intensive reading texts, and short essays for exam practice, thus their exposure to the academic English texts is very limited. Third, the awareness of academic literacies practices seems limited in administrators, teachers and students. This raises questions about the value of the English writing tasks given to these university students for their use of English in their academic and future professional lives.

Lastly, some issues concerning the contextual factors that could influence the English learning and teaching on the research site were not selected for analysis and presented in the study. Still, they can shed light on the role of English, and English language learning and teaching in the inland Chinese higher education institution, which has relatively low levels of international (staff and student) mobility and seems little to be influenced by the general trends towards an internationalisation of higher education.

In conclusion, while there are some limitations to this research, I hope that the learning, use and teaching of formulaic language by Chinese university students in their English written texts presented in the thesis will encourage further steps towards a more nuanced understanding formulaic language, particularly in an EFL context. I hope to be part of the development of a closer relationship between formulaic language research, learning and teaching in the EFL context.

Appendices

Appendix 1 Consent form

Study title: An exploration of formulaic language in Chinese university students' written texts
研究课题：对语块在中国大学生英语写作中的探究

Researcher name: Jiaoyue Chen
研究人：陈皎月
Staff/Student number: 224380547
证件号

	Please initial the box 如果同意请在方框内写Y
I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. 我已阅读此次研究相关信息并有问询权。	☐
I understand that my participation is voluntary and that I am free to withdraw at any time, without my legal right being affected. 我自愿参与此次研究，如果中途退出，不侵犯本人的合法权益。	☐
I agree to take part in this research project and agree for my data to be used for the purpose of this study. 我同意参与此项研究并同意数据用于此次研究。	☐

Data Protection
I understand that information collected about me during my participation in this study will be stored on a password protected computer and that this information will only be used for the purpose of this study. All files containing any personal data will be made anonymous. The data mainly includes:
我知道此次研究搜集的数据涉及到我本人的，将存储在密码保护的电脑里并仅供研究和学习使用。并且所有涉及私人信息的数据将会匿名使用。

	Please tick box 如果同意请在框中写√ Yes	No
I agree to the interview consultation being audio recorded 我同意以录音的形式记录采访。	☐	☐
I agree to the use of writings answers in the publications 我同意我的作文和问卷的答案用于研究出版物中。	☐	☐
I agree to the use of anonymised quotes in publications 我允许匿名援引我的观点。	☐	☐

Name of the participant (print name)……..

研究参与人

Signature of participants……....

签名

Date ……

日期

Appendix 2 Participants information sheet

Participant Information Sheet (Face to Face)
研究参与信息表

Study Title: An exploration of formulaic language in Chinese university students' written texts
研究课题：对语块在中国大学生英语写作中的探究

Researcher: Jiaoyue Chen
研究人：陈皎月

Please read this information carefully before deciding to take part in this research. If you are happy to participate you will be asked to sign a consent form.
在确认参与本项研究之前请认真阅读本文。如果您同意参与本次研究，请在研究知情书页面签名。

What is the research about?
本次研究是关于什么的？
我叫陈皎月，2010 年毕业于湖北汽车工业学院外语系，现就读于英国南安普顿大学人文学院现代语言系应用语言研究中心，博士二年级。此次研究是本人应用语言学博士论文的一部分。南安普顿大学支持本次研究。

这项研究旨在通过调查中国大学生英语写作中遇到的困难，针对这些困难提出解决办法和有效建议以帮助学生提高英文学习能力和英文写作水平。

作为中部教育大省的湖北省，其英语教育在中国来讲既不会太激进，也不会太保守，作为一个本科二批层次的学校，湖北汽车工业学院近些年发展蓬勃，外语系从初步建立到现在的蒸蒸日上，与同类学校相比具备了一定的竞争实力。

每年外语系招收小班制的本科学生，与其他专业的学生相比，外语专业的学生有更多的机会去接触和使用外语。从而搜集的数据会更为丰富。希望在座的同学能够帮助参与到此次的研究中来。对你们表示衷心的感谢！

在研究中，作为参与者将会要求填写两次调查问卷，期初和期末各一次。部分参与者则会有访谈。所以参与者的英语作文将被搜集作为每个人的写作档案。问到的问题基本上都是关于对自己的英语写作的回顾。每次访谈不会超过半小时。总共预计有 7 次访谈。

Are there any benefits in my taking part?
我会从本次研究中得到什么？

没有过多的经济和物质上的，主要是与大家一起分享英语学习经验以及出国学习适宜的咨询。我将会尽自己最大的努力帮助自己的师弟师妹。通过这次研究建档，同学们可以看到自己的写作档案，看一个学期下来自己英文能力和写作上的变化。

Will my participation be confidential?
我的参与是保密的吗？
本次研究符合南安普顿大学数据保护政策（the Data Protection Act/University policy）所有搜集到的数据将不会影响学生期末考试的评测。一旦数据用于发表，将不会使用参与者的真实姓名，通常用替代名或者匿名的方式。

What happens if I change my mind?
如果我中途改变主意不想继续参与研究怎么办？
在通知研究人后可以退出。

What happens if something goes wrong?
如果有意外情况，我还可以和谁联系？
如果有任何意外情况发生，你可以联系研究者学院的负责人（Faculty Ethics Committee）：Prof Rosamond Mitchell (+442380592231, rfm3@soton.ac.uk)

Where can I get more information?
我从哪里可以得到更多的相关信息？
联系人：陈皎月 jc6e10@soton.a.c.uk

研究人导师： Dr. Julia Hüttner (J.Huettner@soton.ac.uk)

更多关于南安普顿大学的信息： (http://www.southampton.ac.uk/)

更多关于南安普顿大学应用语言研究中心的信息 (http://www.southampton.ac.uk/calr/)

再次对您的支持与参与表示衷心的感谢

Appendix 3 First interview schedule (English version)

An exploration of formulaic language in Chinese university students' written texts

Interview guide

Warm-up questions:

Do you like your university? Do you enjoy your life here?

Do you like the subjects you are studying now?

Key questions:

1. Re-assure name, class, hometown, and age of the interviewee.
2. Language learning background: the length of English schooling, extra-curricular English class attendance, stay abroad, learning other foreign language except English?

Content questions

1. Do you think it is important for you to sound like native speaker?
2. Which one do you think it is more important English speaking or writing for a university student? Which one do you think you are better between the two skills?
3. Have you ever pay attention to formulaic language in English in classroom and outside of the classroom?
4. Did your English teachers emphasize on memorizing formulaic language?
5. Do you think formulaic language really help you to be easier to express yourself in English? How?
6. Where do you learn formulaic language?
7. How do you memorize the formulaic language?
8. Do you think your L1 will influence the use of formulaic language in your English writing?
9. Do you think it is easy or difficult to have a good command of formulaic language in English? Why?
10. Do you think highly repeated formulaic language in one text is bad? Why?
11. How do you think about the repeated use of formulaic language among your and your classmates' writings?

Probes

1. Could you point out formulaic language in your writing?
2. Do you draft before writing or just writing without any drafting?
3. How do you deal with the highly repeated formulaic language in writing? Why?
4. Have you tried to avoid the repeated use of formulaic language in one text? How about among class?

The final closing questions

Examples: Is there anything else you would like to add?

Thanks very much for your time!

Appendix 4 First interview schedule (Chinese version)

An exploration of formulaic language in Chinese university students' written texts

预备问题

1.你喜欢你的大学吗？

2.你喜欢你的大学生活吗？

3.你喜欢你的专业吗？

核心问题:

1.核实姓名，班级，家乡和年龄。

2.语言学习背景：在校学习英语时间，是否参加课外补习班，是否在国外生活学习过，是否学习除英语外的其他外语

正式问题

1.你觉得像英语为母语的人那样使用英语重要吗？

2.你认为英语写作和口语哪项对大学生更重要？你觉得自己在两者中哪一项更强一些？

3.你在课堂或者课外有没有注意到程式化语言这种现象？

4.你的英语老师是否强调记忆程式化语言？

5.你认为程式化语言能够帮助使你的英文表达更加到位吗？

6.你在哪里学到程式化语言？

7.你怎么去记忆程式化语言？

8.你认为自己的母语会不会对在英语写作中用的程式化语言有影响？

9.你觉得掌握好英语程式化语言是简单还是困难？为什么？

10.你怎么看高频重复在一篇文章中出现的程式化语言？为什么会有上述的看法？

11.你怎么看高频重复在一个班的作文章的程式化语言？

探究问题

1.能否在你的作文中指出程式化语言？

2.你在写作文之前有提纲吗？或者没有任何形式的提纲直接动笔写？

3.你怎么对待在写作中高频出现的程式化语言？

4.你是否尝试过避免在同篇文章中用一样的程式化的语言？你是否尝试过避免和同班同学使用相同的高频重复的程式化的语言在某次写作中？

尾声

如：还有什么需要补充的吗？

谢谢您的参与！

Appendix 5 Second interview schedule (English version)

1. Do you like English?

2. Do you think it is important for you to learn English well?

3. Are you satisfied with the current learning environment in the university?

(e.g. the equipment in the university, university policy, language teaching and textbook, etc)

4. Are you satisfied with your English learning in general?

(e.g. discuss the learning strategies and results in terms of listening, speaking, reading, writing and translation)

5. Do you think it is possible to use different formulaic language on different occasions? Could you give some examples?

6. According to your learning purposes, please list the following item in the order of importance and give your reason for that.

(Templates, phrases, idioms and fixed expressions, compounds, terminology, phatic expression)

7. Do you have different strategies and purposes on learning different kinds of formulaic language?

8. In general, do you think the language teaching in the university can meet your learning needs in terms of formulaic language? If no, could you specify that?

9. In general, do you think the English textbook in the university can meet your learning needs in terms of formulaic language? If no, could you specify that?

Appendix 6 Second interview schedule (Chinese version)

1.您喜欢英语吗？

2. 学好英语对您重要吗？

3.您对现在的英语学习环境是否满意？

（硬件配套设施，软件， 如：学校政策导向，教师因素，教材选用）

4. 您对现在自己的英语学习总体上是否满意？

（从听说读写译去谈学习方法和效果）

5.您认为在哪些场合会用到不同的程式化语言？

6. 根据您的学习目的，对下列程式化语言安重要性排序

（固定文章模板，短语，俗语及固定搭配，合成词，专业术语，寒暄语/问候语）

7.针对不同的程式化的语言，您的学习策略和目标是否一样？

8.总体上讲，您在的大学中的英语教师的教学，能否满足您对程式化语块学习的需求？如果答案为否，哪些地方您认为需要改进或提高？

9.总体上讲，您所使用的英语教材，能否满足您对程式化语块学习的需求？如果答案为否，哪些地方您认为需要改进或提高？

Appendix 7 Examples of student written texts

Year 1 Student written texts

Task 1: Self-introduction

My name is xxx. A girl come from xxx. I am glad to study in this university and major in English.

Common as I am, I don't have an attractive (attrative) appearance (apperance) or high intellect. Nevertheless, I do everything as hard as I can. What's more, I am a girl who lives a happy life. I always treat things that come through my life as what I have to experience, So that I keep optimistic all my life. Moreover, I have two friends who reach for my hands and touch my heart. They accompany me in a hard time when I really need sunshine and joy.

As I major in English, I want to be translator in the future. In my opinion, a translator's work must deal with much paper and meeting recomde. But I am interested in translation. So that. I think I will do a good job. By the way, I used to have a dream to be a teacher who teach students English. Now, I actually will come true the dream. Be a teacher who can effect our students with great study method and be popular with my students. Maybe I can't be a teacher in the future, but I will always love English and learn English all the rest of my life. I know it's a long way to achieve a goal, but if I always keep on it. I will success. (237 words)

Task 2: My experience

The most impressive experience in my mind is traveling to Xi'an last year. My cousin and I went there with the tour guide.

The first day, we went to Huaqingchi, it was built for Yang Guifei, the wife of an emperor in Tang Dynasty. There was a statue of Yang Guifei, it is so sexy. At first sight of it, some visitors like us felt embarrassed. The tutor guide told us that there would be a show at night, but we didn't have enough time stay there. So it was a pity.

The second day, we climb Mount Hua, it looks like stiff. We have to use our energy to climb the top, it was exhausted. During the road, I met an old man, he said I was brave, I felt a little bit proud to spend two hours climbing on the top. (144 words)

Task 3: The benefit of volunteering

The world exists alone, but every one of us lives in this world cannot exist only by ourselves. We need others' help, also, others need us too. So, it's very important to join the volunteering jobs to contribute our nice and wonderful world.

We may always asked that what can we get from the volunteer job. The answer is, no money, no prize, no reward, even with no thank you. We may thought the answer is darkness, but even it is, there are still many people crowded into the army of the volunteer. These are some typical examples for us.

There is a group of people we cannot even make it clear that how many volunteers at all. The disaster that happened in Sichuan at 2008, because of the destroic earthquake that the rescue car can't arrive at the exact place at one. So where there is a road, there is a group of volunteers go head risked their life and some of them just on their way alone without anything but some rescue food and water. In this war between the human and the disaster, many volunteers even lost their lives with their name unknown to us. They gained nothing even they lost more, but the value of the volunteer will exist at their heart along his life.

And the next volunteering is about me. I undertake a volunteer job at this semester of helping country's students have a class every week. I companied with them each weekend and taught them something out of book. Also, nothing gained but I felt very happy when I do it. The happiness exist in my heart may the most value things from the volunteer job.

The benefits of volunteering can't be accounted by what we get but what we left. The world will be nicer because of the volunteering. (307 words)

Task 4: The Dragon Boat Festival

There are many traditional festivals in China. The Dragon Boat Festival is one of the most traditional festival in my country, and it's also the most unique one, because it is the only festival that to memory a man in my country, and that man is Quyuan, who was a famous poet, and devoted himself to his country in Chinese history. As far as I know, Dragon Boat Festival was held on lunar May fifth every year. It has been held about 2000 years. Since Quyuan was died, it was initially held just in the middle of long river area, and extrended to the whole country later.

Certainly, such a traditional festival, there must have many rituals and activities to observe it. Dragon boat racing is one of the activities. On that day, people would organize many teams to have a boat racing. It was formed because the dragon boat can drive the fish away in the river. So that they can't eat Quyuan's body. Another activity is eating zongzi, which was made from rice and leaves from the bamboo. It's said that the zongzi was to feed fish in the river to avoid the fish to eat the Quyuan's body, and gradually it was eaten by people. It seems that the activities are different from different places, but these two activities is common in our nation.

Dragon Boat Festival has unique meaning to our nation. It is not only a festival but also the ribbon that bond us together to love our homeland. (254 words)

Year 3 Student written texts

Task 1: Should we help strangers?

Recently, there has aroused a hot topic about whether should we help strangers. As series of issues about warm-hearted people who help injured strangers but being defrauded has caught netizen'e eyes. Opinions vary from person to person, but I'm certainly sure that we should help strangers.

On the one hand, people show great sympathy to the well-meaning person. On the other hand, they are wondering whether it is Worthing helping others. In fact, they are just afraid to get in trouble. In my point, I don't think these worrying are appropriate. The reasons are as follows. First, we should be aware that we live in the rule of law country. Prejudice or injustice (unjustice) will certainly punished by law. These criminals will eventually be caught. We live in the democratic (democritic) country. Just remember it. Second, respecting elders and helping others are the traditional virtue. So it should be carried forward all the time. Just give a hand to the strangers and we can warm both hearts. Then, according to the heated topic, we know these unfair events are special cases. It is rare. The small parts cannot refer to the whole country. It's no deny that justice is powerful than crime. Helping others is the good thing and most of the people are surely proud of you. At last, tomorrow will be better. This tells us beautiful things are on the way.

All in all, one for all, all for me. Today, you give a hand to a stranger; in the future, another stranger may help you when you are in trouble. So, I really agree to help the strangers bravely. (272 words)

Task 2: Qing Ming Festival

In the spring and autumn period of China's history, Duke Xiao was the monarch of the state of Jin. One of the King's sons was called Chong Er. Jealous of his talent, a concubine falsely accused him of rebellion to make her son the crown prince. He had no choice but to flee and with his mere some officials. They hid themselves in a mountain and went hungry for quite some time. An official name Jie Zitui took great pain to cut some flesh from his thigh and cooked it for Chong Er. When the fact was known the young master was moved to tears and knelt down in gratitude. And Jie replied his best repayment should be a just king. They lived a life of hunger and cold for three years until the evil concubine died. Many soldiers were sent to look for him and to escort him back home. Going into the carriage, he saw an official packed on an old mat onto a horse, he said laughingly, "What on earth is the use of that? Throw is away!" Jie Zitui heard it and sighed, "it is hardship that can be shared with his majesty but not prosperity!" So he went away quietly and lived in seclusion with his old mother.

As Chong Er became a king, he rewarded many people but he forgot Jie Zitui. He did not realize (realise) it until was reminded. However, his invitation was refused and he flared up. Soldiers were ordered to burn up the mountain to force the Jie to come out. Finally, they found Jie and his mother scorched under a willow. He would rather die than yield to the power. Chong Er was so overwhelmed with regret that he ordered people hold memorial ceremony for Jie. So every year on that day, folks mourned for him and the day before the cold meals, which avoided making fire. Later, the custom of inserting willow branches on gates was also added. (330 words)

Task 3: A book / movie review

A year ago, a movie named "Love Actually" impressed me most. This English movie is made up of several short stories, but the characters in the stories are related. There are romantic stories between Prime Minister and a new junior member of the household staff, there is a kind of silent love between a writer and his housekeeper, there is also a love between the father and the son, we can find actual love with different meanings in the world, every meaning of love will tell us a moving story and give us hope as well as warm.

The film begins with a voiceover from David commenting that whenever he gets along with the state the world he think about the arrivals terminal at Heathrow Airport, and the pure uncomplicated love felt as friends and families welcome their arriving loved ones. David's voiceover also relates that all the messages left by the people who died on the 9/11 planes were messages of love and not hate. The film then tells the love stories of many people.

Focus on the core of topic, love actually, it picks the background the time of Christmas period, all romantic stories of the curtain slowly in order. Every love story gets it happy end, every kind of love is spreading to every corner of the earth, on the screen, more love scenes appear and together they coverage to one----actually love is all around. (239 words)

Task 4: The importance of reading classics

Recently, the issue on whether it is important to read classics has aroused hot attention in public. Some people hold the sentiments that practical operation is more important than reading classics. What's more, some others think we can get enormous information on TV or

internet. There is no need to read classics. But we have to respect the fact that reading classics attach great importance to people's grow up.

We have noticed a phenomenon that there is a decreasing number of people who are ready to read classics. For the reason that most of people haven't developed a great habit of reading. With the temptation of mixed and complicated information provide on TV and Internet, more people trend to obtain knowledge from TV or Internet instead of reading books. What's more, the information on TV or Internet have both good and bad news, we have difficulties in choosing (choicing) or absorbing fabulous information. On the other hand, some do have habit of reading, but what they read are less worth reading. They read novels or essays which have little nutrition, so they get nothing after reading and their efforts are in vain. Others have the ideas of playing, watching TV or some other else. They are just lazy to read and don't have concept of enhancing their inherent qualities.

In order to solve the serious problem, some actions have to be taken in our lives. First, we have to build up a concept of reading classics and develop a good reading habit. Second, we have to know how to resist the temptation from the outside world. Lastly, we should love reading classics instead of considering it as a burden or duty. (281 words)

Appendix 8 Interview examples

Example 1: Frist round of interviews

(INV-P1-001-M)

A: 首先我得说，这篇作文并不能代表我的作文水平，因为，它的篇幅有限。

Q: 没关系，这个采访跟你的作文水平没关系，跟你的作文的字数也没有关系。然后跟你的语言对不对错不错也没有关系。就只是描述，不做判断。所以，不用考虑成绩。

A: 好的。

Q: 你想用中文还是英文？

A: 都可以。

Q: 那先中文吧，下次等适应了我们在慢慢用英文。杨添，也是一班的。

A: 是的，没错。

Q: 然后家里是？

A: 我的祖籍是荆门的，现居在十堰。

Q: 今年几岁呀？

A: 18 岁。

Q: 这么年轻！

A: 嘿嘿。

Q: 然后什么时候开始在学校里面有英语这门课的？

A: 小学三年级。

Q: 有没有参加过任何的这种课外的补习班？

A: 在小学一年级的时候就开始参加一个入门的…

Q: 像泡泡少儿英语的, 就相当于那些？

A: 对，没错。然后在小学的时候就开始学那个新概念第一套，对我初中学习很有利。

Q: Is that your handbag?

A: Yes. Which made me good grades in the high school. 当然在初中的时候我也曾想过继续学习新概念，因为我觉得，确实是非常有用。但是由于时间等方面的限制，我没有跟班学习，而是选择自学。但是在英语学习上，我很少遇到问题，因为很巧妙的一些难点，都被我不经意的克服了。比如说，我当时并不知道单词的构架作用这么重要，但是，我就是记单词，当然这对我的成绩很有帮助。英语，作为几个主科里面， 我还是比较放心的，但是我发现，反而进入高中，我的英语成绩反而没有我在初中和小学的英语成绩那么理想。

Q： 为什么呢？你有没有想过原因呢？

A: 我当然想过原因，当然想过这个问题，当然，首先这跟我不是很熟悉高中那种学习节奏，然后它没有那种语言环境，其次，一切都是立足于考试的。

Q: 那你初中小学就不是嘛？

A: 不是。初中的时候他们强调说，它们根据语言，社交活动和这些用语，但是，因为，英语作为一门语言，它的第一用途，还是用于作为交流，当然，进入高中之后，它的作用显得有点奇怪。我反而感觉成绩 有点下降，当然它也不是很差，一般考试分数会在 120-130 左右，当然，毕业考试英语是我最高的一门学科。

Q: 然后你之前有没有出过国？

A: 没有。但是我有这样一个打算，当然是要选择留学，选择考研或者考国外的大学。这是在考虑范围之内的，之前虽然没有出过国，但是我父母经常为我创造更好的语言环境，经常参加夏令营，而且我认识一帮口语非常棒的人，所以，我很乐意与他们交流。

Q: 那这种夏令营是属于英语夏令营，与英语的 native speaker 在一起的还是 with high proficiency level 的这种？
A: It's a combine. It's combining the two level, make conversation, and the lower level can have progress. It's very beneficial.
Q: So, it's mainly more oral or speaking stuff not the writing one?
A: Yes, right. Yes, it's oral English practice and very important, as I am concerned. However, now, and all the time I find in China, they haven't any very good oral English speaking environment. I try my best to make conversation with foreigners, like our foreign teacher. I made conversation with them, it's ah, although, my oral English is not as fluent as them, but I found that I'm better than my classmates.
Q: Yeah, exactly.
A: They said that why you have, you mean to train it all the time, I said no, I just make a conversation with others. And then they said my English grade is better than you, but it is not my oral English. I said, because you pretend to agree to writing, reading, listening, but no speaking. They said, ah, I see.
Q: Do you think you communicate with native speaker or the people with higher proficiency level in English; is that a factor which makes you better in speaking than any others in your classmates?
A: Yes, I think so. Well, I also think that if you do not speak fluently, or constantly, you cannot speak fluently, that makes me to have better oral English than my classmates. And I am also concerned about my English writing, listening and reading, however, because I got stuck with some affairs, which makes me didn't have good grade in the second semester.
Q: So, primarily, you think speaking is the most important one among the four skills?
A: In fact, English as a language, its first usage is to make a conversation for us to communicate. However, as a language, it also has other usage. Well, the communication is the most important of course as all the language. But however, if you an English major, you would better to do better job at writing and other kinds of things than others, also is very important.
Q: Do you like writing? I mean, in Chinese.
A: Of course. As in Chinese, I keep writing in articles, like dairies, essays, and all kinds. I pretended to publish a book for myself, but it is still far away from this. And my teacher Daniel often said that, "you are an enthusiasm writer, however your writing is lack of idiomatical convey and what you should do is read more articles written by the native English speakers, and it's also very important to learn some words by heart and some idiomatical usage learned by heart." And of course I'm trying to do it and I have ever made a better job in last semester.
Q: So, which one do you think you are better at? Which one you like more compare writing in Chinese and writing in English?
A: Of course writing in Chinese, because, there is some usage that you cannot convey in English. For example, there are a lot of models or word strings you cannot translate into English idiomatically, so. Also they are some poems are considered beautiful in Chinese where we translate into English, nothing. Also, there are some poems in English written by Shelley or someone else, and if you translate into Chinese, uh…
Q: Sounds weird.
A: Yes.
Q: Do you like writing in English?

A: In English, I am also an enthusiasm writer, but however, I'm lack of skills. I'm very lack of skills in writing English articles, comparing with my Chinese articles. I think I need to improve idiomatical usage in English, and how to write really English article.
Q: So, in your words, "idiomatical" means native alike?
A: Yes.
Q: Do you learn other foreign languages currently?
A: I learned a little bit Japanese and German. German is recently learned.
Q: So you are a quite beginner?
A: Yes, a quite beginner.
Q: Alright. And this research is about formulaic language, as I said, it's a kind of language you used as a whole, as a chunk, as prefabricated, and you learned that, you knew that, you met that, you memorized that, you used that together, without analysing. Other than you use some language, you think about the grammar of that. So, have you noticed this language phenomenon before?
A: Yes. I'm quite lucky because my teacher said I have a gift in language. I have thought about some grammar phenomenon, when first connected with English, while something is quite like Chinese, while other things are quite ridiculous, pretty strange. However, when you learn more, you found that English grammar as well as Chinese grammar, there is a just rule in language, you can obey this rule in writing, well, you didn't care too much when you in oral English. Because in writing, you write something down on paper, thought it requires everything is right, and cannot make any mistake by others. So it requires correct grammar. While when you are speaking, if you concerned too much on the grammar, well, it will stuck you from making conversation with others.
Q: So, there is time limit?
A: yes, time limit. While others, you will concern more about others' mistakes rather than what's he said. So that's why I said grammar is important in writing rather than in oral English. Also the grammar, I made the comparison, from the grammar for Chinese, English, and Japanese. I think there are different systems in different grammar, as well as in Japanese and Chinese grammar, they are in the same system, while the German, English and Latin, there are other systems.
Q: From other language family.
A: I think I can use the former grammar system very; I used it as if it's just in my hand, but while you used other systems you have to make a switch in the hand.
Q: Have you met so-called formulaic language in your English learning, in your middle school, in your senior high school?
A: Yes.
Q: Even in the university? Can you just give me some examples for that?
A: I beg your pardon?
Q: Have you paid attention to the formulaic language in your language learning, in your English language learning before?
A: You said in the primary school?
Q: No, I mean, from the middle school till now.
A: Till now? Yes, indeed. For example, I will be taking writing as an example, if I write an article; I write and check three times after I write it. Because I write article very enthusiasm and take few minutes to finish it. But then, I will check it out, whether there is grammar mistake or not, it is

the second; the third time I will check some usage is not idiomatical or wrong, I will refer to dictionary, refer to teacher, to check it out and make sure I won't make mistake another time. While I'm speaking, sometimes I'll be thinking whether this grammar usage is right, or this is a mistake. And learning the grammar use strategies where I make conversation with foreigners. When I asked them how I can use the English grammar, and make less mistakes, and they said, "the only way is when you speak, you needn't care about grammar. While you read articles, or practice writings, you pay more attention to grammar." This is maybe a useful way; I don't know how to say it.
Q: Do you think it is important to use English like an English native speaker?
A: That's very important. As a language learner, our English major, a language for you is a tool, to communicate with others. If you learn it just as some academic, you didn't use it. After three years later, ten years later, you cannot remember a single word.
Q: Do you think to make someone understand your English is more important than you speak like a native?
A: Yes, of course.
(A's alarm screaming)
Q: That's fine, no problem. Do you think to make someone understand your English is more important than you speak like a native? Which of them is more important? Or you think both of them are equally important?
A: Of course, as a tool, to make somebody understand me is the first thing that I have to concern. For example, you are English major, if you cannot make others fully understand you, what's the meaning to you to learn this language？
Q: You think the communication comes first.
A: Yes.
Q: Alright, speak to the formulaic language, it refers to some kind of fixed expressions, certain kinds of writing structures, and some idioms, greetings, such as "How are you?", "Fine, thank you, and you?" even someone's nickname, someone's fixed expressions related to someone's typical personalities, have you paid attention to this before?
A: You know, I love English film, while in the films I paid attention on the models, some idioms, and so on. Firstly, your saying is interesting; you have interest in it, so you started to learn about it, and then when you master it you can use it in your life. Now I still remember that when I first front something very English style, I made a comparison. I often make comparison with Chinese and other languages, those idioms are some nicknames, they have something equal in Chinese, it is not so hard to understand.
Q: Yeah.
A: Yes, there is, for example, you can call Peter PITE, or something, they can fully understand in Chinese some nicknames. So that's connected. There is nothing different. But if you learn more deeply, you found that in English, they have very funny phenomena, when you use idiom. Because English has developed thousands years procedure, in different procedures it is changed a lot, up till now, English is quite different from the begin, and different words wording to combination with other language, well, that I really I have a long way to go. I need to learn more about idioms.
Q: Do you think the formulaic language is important for your language learning?
A: I think so.

Q: In what kind of ways?
A: Well, it's pretty hard to say,
Q: Or you can speak in Chinese that's fine.
A: Alright, that's quite strange.
Q: No. it's fine.
A: 但是我觉得，如果说英语的是，我们学习英语首要的是为了什么，首先，我觉得我学的英语专业，它不算是一个专业。学习英语的人大概抱有两种目的，一种人认为，我是要研究英语，并且是文学的； 当然，另外一种人是把英语当做一种工具，需要其他的专业知识。 我觉得我自己是属于后者，当然使用英语作为交流语言当然是它作为工具的第一用途，所以我觉得这当然是最重要的了。就我看来，当然，如果是以英美文学来讲的话，去深入的了解这门语言它的历史，它的用法，它的优美，并且能欣赏他是第一用途。当然，我也爱好文学，但是我自觉得研究英美文学，我是作为一种兴趣爱好，而不是专业。我会把语言作为一种交流工具，更看重这一点。
Q: 那你你觉得我说的这些语块，这些固定搭配固定表达，包括一些谚语成语俗语这些，你觉得重要吗？它们扮演一种什么角色？
A: 重要。首先，这些东西是需要逐步积累的，就是通过你的读书。
Q: 为什么要积累？
A: 因为如果你不能把这些融入到你的英语学习当中，那你学的是什么？就像中文，你没有学习古文，没有学习诗词，没有学习语法的话，你的中文会非常糟糕。但这应该是一个语言生动活泼所在，一门语言的特色所在，也是一门语言生动活泼的地方所在。也是我感兴趣的，所以说，这当然很重要。
Q: 就是使你的语言更加有张力。然后就是更地道。
A: 是的。
Q: 那你觉得这些东西是非学不可的嘛？用你的话说就是我学外语就是必须得学习它的固定搭配固定用法。
A: 是这样的。通过这样的学习一门语言，你才能更好的了解为什么这门语言成为一个世界的交流工具。 没有人考虑这个问题，我在考虑这个问题，首先是因为它简便，其次他能把很多有趣的事情用很简单的方式表达出来。
Q: 比如说？怎么能把有趣的事情简单的表达出来？
A: 在中文里面，如果描述一个东西，它需要一种描述手法；但在英文里面不用。比如说，don't on my way. 用中文想就是不要站在路上，就要转化成"挡着道儿了"，但是英语就不需要。On my way 就是挡着道儿了。它不要一个转换思想的过程，就像一个很形象的事情就很容易用英文表达出来，但是中文不行。
Q: 那你觉得就是说你的小学，中学，甚至大学老师会不会强调这些固定的语块词组以及模板的用法？还是会有所变化？从开始的基础教育，到中学的应试教育，到大学的英语专业？在这些课程中你的老师对这些语言现象会不会有所变化？
A: 当然，当你以 exam 为目的的时候，老师们肯定会更注重高考所需要的这一类的语言的用法。比如说，文章的连接词，再比如说一些好的开头所需要的一些用法，再比如说催人向上的一些谚语，他们会看重这些在英语写作中的好处。他们在考试中也会很注意考试的阅读，你怎么把它看懂，所以说他们也会提倡这个。但是你到了大学，没有人追着你屁股后面让你学这些东西，完全是出自 Academic 的一种爱好，或者是一种非常学术的，它为什么会这样。
Q: 就是靠自己的比较多。
A: 没错。

Q: 那你现在还是有留意这方面的积累嘛？
A: 有。
Q: 那你通常都是怎么去积累的？能不嫩给介绍一下你的方法经验，通过什么样的手段去学习？通过什么方法去刻意地记住这些？
A: 阅读。首先是阅读，因为文字给你的直观感受在书面上会比较强烈，比较明显。你可以随意地把它圈下来，写下来，记下来。把它的用法记下来。这就是最简单的方法。当然，我会采用其他方式，比如，当与一个外国人交流的时候，他用的这个用法我好像从来没有听过，我会问一下，我会问下这是怎么用的。为什么要这么用。交流是其次。当然，我觉得最重要的还是要读书。读英语原著。
Q: 还是以自己学习为主。学完之后有一定的积累，再跟别人交流，然后再去学习。
A: 嗯。
Q: 就是说现在这种语言现象中会有一些高频词汇，出现在你的作文中，出现的次数比较多，比如第一段第二段你都出现过，你觉得这是一件坏事吗？
A: 文似看山不喜平嘛，当然你要用词重复的话，显然这是一件坏事。
Q: 然后你觉得会对你有所影响吗？
A: 首先我要考虑的是别人对这篇文章的一些看法，因为你作为 English major, 你既然能用同一个单词用两回，我觉得应该是不可饶恕的。
Q: 如果这些词出现在你和你同伴的作文中你会担忧吗？
A: 会。
Q: 会不会担心影响自己的成绩？
A: 除此之外，我还会非常担心他们会怎么写，他们会怎么写我就一般就绕开它。
Q: 那你都有什么策略去绕开这两种情况呢？
A: 这些策略在书里面，比如说，有些英国人写的文章，或者是他们写的书，他们会有很多很多类似的用法。你不必要非要是课本上说的那种，或者是学到的那种。你要结合自己课下学到的来避让这些雷同的地方。
Q: 具体的方法呢？是同义替换还是同个意思换不同的句式？
A: 当然你要表达出同一个意思不一定要用同一种用法，除了你提到的，还能换个模板来说。
Q: 最后一个问题，就是你写作文会有提纲吗？
A: 没有。从来没有。
Q: 心里想的，腹稿，不是写在纸上的。
A: 一般我看到题目的时候我就知道要写什么。
Q: 那“写什么”这些话一般是中文的还是英文的一般？
A: 我还没有达到这个水平，一般都是中文的。因为思维方式不能达到立马就好。

Example 2: Second round of interviews
(INV-P2-007)
Q: 刘光洲，你喜欢英语吗？
A: 挺喜欢的。
Q: 那你喜欢学英语吗？
A: 喜欢学英语。
Q: 以现在这种方式吗？
A: 嗯。

Q: 那么你有没有想过理想状态下学英语是什么样子的呢？
A: 我可以说就是我们现在主要上课，比方说，先上的是英美文学这种，然后像高英那种，然后到最后就是以考试来衡量你学习的成绩。但我觉得考试这种形式就太无聊了。首先是它一方面考查的不是你的学习能力，另一方面，还有语言它本身，它需要输出，最终是要说出来的。但我们现在的考试根本就没有，它就是考
Q: 记忆力？
A: 对对对。我觉得这种考查你能力或者说语言输出的能力，应该不能完全靠书面这种考试来衡量来检查。
Q: 那你就是说这个学习结果的评估，比较死板？
A: 对对对。
Q: 那么你理想中的学英语的方式应该怎么去学呢？
A: 我觉得大家应该更侧重于说吧。然后说和写这方面，还有听。
Q: 对于现在学校这种学英语的外部环境你有什么看法吗？你觉得对你的英语学习有什么影响？满意吗？
A: 我觉得不满意。首先，咱们学校是一个工科院校，以工科为主。大家都知道我们学校每年的四级通过率也不是太高，还有很多人到毕业的时候还过不了。很多人在平时根本就不学英语。
Q: 那为什么一定要学英语呢？
A: 因为现在是经济全球化吗，不管你到哪儿， 英语都能用得着啊。
Q: 但说中文的人也很多啊。
A: 但是你像现在很多核心的技术都是学外国的，或者说很多设备上全部都标的是英语。如果说你不会的话，就算你是工科生，技术上你都懂但是你看不懂上面标的是是什么。我觉得如果你想在某一个领域想做精的话，我觉得会很难。
Q: 那他们都会了的话那我们学这个专业就不用开了啊？
A: 咱们文科的话虽然是只学语言，但是咱们以后可以出去做外贸这块，或者说，教外国人汉语，就是语言交流这块。而不是说技术那块。
Q: 那你觉得就是说学校不太重视，那还有没有别的方面呢？包括老师啊，你用的教材啊。
A: 我觉得我们现在用的教材还好吧，我也看过，很多学校跟我们用的教材一样。像高英，大部分外语院校都用的是张汉熙的这一版。然后还有英美文学，也有的学校要是考研的话也用的是那个版本的。我觉得教材都差不多的，都挺好的。关键就是老师上课的时候，有的时候就比较抢。你比如说那个高级英语，就很多就需要花很多时间去品味那个文章，不能够就是老师在哪儿讲修辞手法还有那个句型什么的，我觉得那个东西如果你感兴趣的话，你去从里面慢慢，我可以就是说花个十来分钟就是去品析一个句子。我就算不知道它是一个修辞，我把词的意思查一下，然后慢慢地多读几遍，我觉得我自己多读几遍也能理解。如果说老师的强制灌输，这是什么修辞手法，然后这个句型要记住，然后怎么怎么着，我觉得这样的话学的太死板了。大家也就没兴趣去记这些东西。所以我觉得上课方式，我觉得对我自己来说，我觉得我不喜欢这样子。
Q: 那你理想中的上课方式是什么样的呢？
A: 老师可以把课堂交给我们叫我们自己来读，上课可以给一定的时间，自己读自己理解。老师就是引导一下就行了。因为语言这个东西，本来就靠自己的听啊说啊写啊。
Q: 但是有些学生很差呀，你不强制灌输他就什么都不知道啊。
A: 那老师可以引导他们，就是说从思想上改变他们，叫他们多花时间嘛。
Q: 也就是说给学生分个层次。
A: 我觉得这么可以，工科班已经分了。

Q: 那我问一下你觉得学习好英语对你来讲重要吗？
A: 挺重要的。
Q: 重要在哪儿?
A: 首先我觉得我现在是英语专业，我以后就业方面就是说肯定与英语有关吧。然后第二个就是说，对于那些喜欢旅游的人来说，就现在出国旅游也很频繁，如果他想出国的话，在语言这块就不需要太依赖别人，比如说导游哇，或者说朋友跟着一起去呀，可以一个人想到哪儿就到哪儿。因为英语基本上是每个国家都会的。
Q: 世界性的？
A: 对呀。
Q: 那你现在对自己的英语学习总体上是否满意？
A: 不咋满意。
Q: 为什么?
A: 因为首先就是现在这种以书面形式来考核你的学习能力的话，我觉得我书面的解答答题能力不是太高。第二个我感觉自己的听力和表达能力也不是太好。我觉得一方面是因为我们学校缺乏这样的一种环境。比如说好多内容就外国语院校就他们去考试，像那种需要播放广播呀新闻啊要用全英文的，然后还有就是像语言类这种，他们的专业比较多嘛，大家很多时候有很多机会就可以说，但是在我们学校就不行了。所以说在表达这块，有时候你想说吧，也没人跟你说。一个人说也没啥意思。还有听力，就是我觉得我们用的那个教材不咋好。不过现在改版了，估计改的好一些了。我们买了 4 本书，但就听了两本书，还有两本书没听。
Q: 为什么？
A: 不知道，上不完吧。因为有专四强化训练，所以耽误了。写作课还有口语课的书全是新的。没用过。
Q: 那你觉得你的写作课对你有没有用？
A: 有点帮助吧。
Q: 比如呢？
A: 比方说写作格式那些什么东西啊。如果老师不讲的话，原来根本不知道。特别是像商务信函书信啊啥的。虽然是学的的确比较少，但是我觉得还是学到了一些东西的。
Q: 那么根据我纸上列的这些程式化的语言，请你按自己对它们的重视程度排个序，然后告诉我你为什么这样排。
……
[排序中]
……
A: 我觉得排序的话首先就是专业术语，然后是口头禅问候语寒暄语，然后是俗语和固定搭配，然后是短语，然后是合成词，然后是固定文章模板。我这么排的主要原因就是我觉得如果，你说英语，你表达首先是要别人能听懂，你的表达不能出问题，所以说专业术语的话，不管你在哪个领域，对，你说的是专业术语，这样的话既方便别人能听懂，也更有利于你跟别人交流。然后再下来就是专业术语口头禅还有就是那个俗语和固定搭配这三个，我觉得都差不多吧，就是说能更方便跟人交流。然后后面这个，短语，固定文章模版和合成词，我觉得记这些东西，在一定程度上在书写方面，能够会为文章增色。因为像短语，固定文章模版，这个首先就是格式，你如果按照那样写的话，让人看起来会感觉很清晰。然后就是短语和合成词，比如说有些很新鲜的短语，首先是别人看到这个短语很容易能够读懂，第二个就是，看的比较新颖。因为平常有些短语大家不用，如果你用在文章中，看起来比较新颖。

Q: 那么根据这些不同种类的程式化的语言，那么你觉得在你学习的过程中所用方法，策略还有你期望达到的目标，比如你刚说的，专业术语一定要精准，有什么不一样的要求吗？
A: 我觉得这么多的语言，咱们作为一种外语学习吧，我觉得最重要的就是背然后就是说，就是这两方面。因为我们有很多东西，比如像专业术语和口头禅，在我们接触之前我们可能根本就不知道。要想会用肯定要先把它们记住，然后加以灵活的运用，然后才能慢慢把它们掌握吧。
Q: 那你觉得你的大学英语老师在课堂上的教学能不能满足你对程式化语言学习的需求？
A: 大部分的老师还是提到的比较多。比方说短语啊，固定文章模板，还有俗语口头禅。
Q: 那么老师是提到了这些语言，那老师的教学有没有介绍一些学习程式化语言的一些方法？或者说他们在这方面的教学是怎么促进你的学习的？
A: 有的。
Q: 怎么样的?
A: 像固定文章模板，之前我们考专四的时候或者像原来高中老师说过像什么三段式啊，就是写议论文啊，还有像短语啊什么的，这就是必须得记得。还有口头禅问候语寒暄语啊这些，老师叫我们背对话。就穿插在文章中。还有就是合成词和专业术语，这些老师就是让我们死记硬背。
Q: 那么总体来讲你的大学教材对这方面是否有所强调？比如说，大学教材上的内容能不能满足你对英语语块学习的需求？
A: 我觉得不够。比如说我们现在学高英，它编写这个教材，就是涉及到各个方面，像什么文学类啊，还有，主要是文学类的。它就是像什么很多生活化的语言它那个里面就涉及不到，还有像固定文章模板啊在里面也没有反映出来。或者是合成词啥的，比较少。然后还有就是，就是我们现在用的教材，上面这六个方面，不能通过对教材的学习把这六个方面都掌握。
Q: 那么主要能反应哪些方面呢？
A: 我觉得最能反应的是短语，专业术语，口头禅，还有俗语及固定搭配。差不多就是这四个吧。但它主要是部分的。
Q: 也就是总体上它的这个内容还不能满足你的学习需求?
A: 差不多。
Q: 好的。以上是我的所有的问题，你有什么问题吗？

Appendix 9 Coding categories

- **The coding categories for the first round interviews**

Understanding formulaic language

- Form
- Function

Formulaic language learning

- General learning process
- General learning strategies
- General sources used for learning

Using formulaic language

- Process of using formulaic language
- Strategies in using formulaic language

Formulaic language teaching

- Foreign teachers
- Domestic teachers

- **The coding categories for the second round interviews**

Understanding six types of formulaic language

- Writing templates
- Idioms and fixed expressions
- Phrases
- Phatic expressions
- Compound words
- Terminologies

Learning preference of the six type of formulaic language

Learning strategies used in six types of formulaic language

- Writing templates

- Idioms and fixed expressions
- Phrases
- Phatic expressions
- Compound words
- Terminologies

Students' comments on formulaic language teaching

- Teaching methods
- Content
- Textbooks

Learning contextual analysis

- (The internal elements)
  - English language learning motivation
  - Reflections on the English language learning
- (The external elements)
  - The university policy
  - The teaching staff in the university
  - Peers

Appendix 10 Student identified formulaic strings

Year 1 student identified formulaic strings

No.	strings	Notes
1	As well	
2	…be drawn	
3	a freshman majoring in	
4	A kind of	
5	a money mighty world	
6	A strong feeling of being	
7	A three-day holiday	
8	A very place where hold	
9	According to	
10	achieve my dreams	
11	After all	
12	After graduation	
13	After that	
14	All kinds of	
15	And so forth	
16	Are fulfilled	
17	As a Chinese	
18	As a result	
19	As an old saying goes	
20	As before	
21	As far as I am concerned	
22	As for	
23	As is known to all	
24	As it is known to us all	
25	As it is said	
26	As their tradition	
27	As to	
28	As we all know	
29	At that time	
30	At the same time	
31	At the side of	
32	Baby doves	
33	Be admitted as	
34	Be faced with	
35	Be filled up with	
36	Be given	
37	Be responsible for	
38	be so honour to	
39	Because of	
40	Been asking for perfect	
41	Benefit from it	

42	Best occupation	
43	Bring…up	
44	Build up my body	
45	By the way	
46	Come to our mind	
47	Comparing with	
48	Consist of	
49	Devoted himself	
50	did well in	
51	Die for	
52	Drive evil and ghost away	
53	Earn money to support	
54	Even though	
55	Family quality time	
56	Finish it well	
57	First…second…last but not least	
58	Firstly...Secondly…Finally	
59	For example	
60	For instance	
61	For one thing…for another	
62	From my point of view	
63	from then on	
64	Gain something good from it	
65	get along very well with	
66	Get home	
67	Get something useful	
68	Get together	
69	Get up very early	
70	Give me a deep impression on	
71	Give their hand to	
72	Go away	
73	Go back to	
74	Go to the river	
75	Had no time	
76	Have trouble communicating	
77	Having delivered myself about	
78	High ranks	
79	How beautiful the scenery is	
80	How exciting	
81	How exciting game it is	
82	Hurried to	
83	I am behalf of…	
84	I am good at	
85	I believe that	

86	I hold the view that	
87	I hope I'll	
88	I think it's a	
89	I was admitted to	
90	I'll try my best to be impressive	
91	I'm not skilled at	
92	If permitted	
93	If you want to be loved, love others.	
94	If you…you will	
95	If...I'll be	
96	Improve your skills	
97	In a word	
98	In addition	
99	In conclusion	
100	In fact	
101	in honour of	
102	In lunar	
103	In order to	
104	In other words	
105	In recent years	
106	In the ancient China	
107	In the meantime	
108	In the morning	
109	In the river	
110	Is far away from…	
111	It is a special day for us to	
112	it is my pleasure to be	
113	It is necessary for	
114	It was time for me	
115	It would be much happier of giving instead of receiving	
116	It's said that you can judge a man from his hobbits	
117	Join in	
118	Jump into	
119	Keep…from	
120	Know what happened around	
121	learn a lot from	
122	Learn…well	
123	Less and less	
124	Look upon	
125	Lose weight	
126	Made great contribution to	
127	Make my dream come true	
128	Make new friends with	
129	Make…into	

130	Maybe you know	
131	More than	
132	Most of	
133	Much more… than…	
134	Nice to meet you	
135	No matter how	
136	no pain, no gain	
137	Not only…but also	
138	Now that	
139	Of course	
140	On earth	
141	On the one hand…on the other hand	
142	On the other hand	
143	One of the	
144	one's loyalty to	
145	Open my eyes	
146	Out of kindness	
147	Out of trouble	
148	Pay attention to	
149	Prefer to	
150	Primary school	
151	Protect…from	
152	Put it differently	
153	Put up	
154	Put…into	
155	Put…into practice	
156	Realise dreams	
157	Sacrifice…for	
158	Six years later	
159	So great to the ear	
160	So…that…	
161	Some of	
162	Something about it	
163	Something helpful	
164	Speaking the…	
165	Such as	
166	Support my family	
167	Surprise may be,	
168	Take a trip in honour of	
169	Take care of	
170	Take part in	
171	Talk about	
172	Talk with	
173	Taste sweet in heart	

174	Tend to be	
175	That would be	
176	that's mean	
177	That's to say	
178	The air is also fresh for us to breathe	
179	The bones of fish	
180	The delight of giving	
181	The Dragon boat festival	
182	The food is to Chinese people what the turkey to the Americans	
183	There are	
184	There might be	
185	Think of	
186	This is to say	
187	Those who	
188	those who are needed help	
189	To affect everyone	
190	To be a…is	
191	To begin with	
192	To memory this great person	
193	Traditional festivals	
194	Try our best	
195	universal love	
196	Used to be	
197	Very much	
198	Want to honour	
199	What a bad luck!	
200	What a busy day	
201	What a lovely day	
202	What about my dreams?	
203	What you have learn	
204	What's more	
205	When it comes to	
206	Where there is river, there is a city	
207	Which was said to	
208	With time going by	
209	You are due to	
210	you can call me	
	Though	Individual word
	But	Individual word
	Nevertheless	Individual word
	And	Individual word
	Meanwhile	Individual word
	Besides	Individual word
	Gratification	Individual word

	Unfortunately	Individual word
	Outstanding	Individual word
	Admirable	Individual word
	Value	Individual word
	As a result	Overlapped
	As far as I am concerned	Overlapped
	As is known to all	Overlapped
	At that time	Overlapped
	At the same time	Overlapped
	From then on	Overlapped
	In addition	Overlapped
	In order to	Overlapped
	In the meantime	Overlapped
	No pain, no gain.	Overlapped
	Try our best	Overlapped
	When it comes to	Overlapped
	Having delivering myself about…	Subsumed
	How exciting	Subsumed
	I hope	Subsumed
	Not only…but …	Subsumed
	Prefer to	Subsumed

Year 3 student identified formulaic strings

No.	**strings**	**notes**
1	More important	
2	Not familiar with	
3	First and foremost	
4	Not long ago	
5	Getting fewer and fewer	
6	First of all	
7	Last but not least	
8	Later on	
9	More and more	
10	More or less	
11	Rather than	
12	Sooner or later	
13	The last but not least	
14	Not all	
15	I firmly believe that we	
16	One never knows that	
17	We must be clear that	
18	As it is known	
19	We, as college students, should be	

20	Not only…but also…	
21	Only when…can we	
22	Lucky enough to	
23	Try one's best to	
24	See what we are doing	
25	what attracted each other maybe	
26	What impressed me most	
27	What you should do is to	
28	which is	
29	which strongly illustrated…	
30	Pure Brightness festival	
31	Qing Ming Festival	
32	Tomb sweeping day	
33	Dreams of becoming	
34	Lack of	
35	One of the most important	
36	Return of colour	
37	The passing from one generation to another	
38	Those in need	
39	A three days off	
40	Ancestral tablets	
41	Heartless and soulless	
42	Jin Wen Gong	
43	The coming of an era of	
44	Unexpected twists and switchbacks	
45	Various choices	
46	Waste-time and waste-energy thing.	
47	As far as I am concerned	
48	Head over heels	
49	To help or not to help	
50	Turn a blind eye to	
51	Turn a deaf ear to	
52	What's more	
53	No pain, no gain	
54	Do onto other as we would have them do onto us	
55	Help others is help ourselves	
56	Greed is the eternal soil of violence	
57	Like two stars, crashed together	
58	Only the deaf appreciate rearing, only the blind realise the manifold blessing that life in sight	
59	Rose to lover, smell in hands	
60	The roses in her hands, the flavour in mine.	
61	Where there is a book, there is gold	
62	As well as	
63	Few and far between	

64	First…Second…Finally	
65	First…Then…	
66	To begin with…then…at last	
67	Hot issues vary from day to day	
68	Above all	
69	According to legends	
70	All in all	
71	For example	
72	In order to	
73	In spite of	
74	In terms of	
75	Instead of	
76	At last	
77	At the same time	
78	On the other hand	
79	For thousands of hundred years	
80	From my point of view	
81	In my point of view	
82	Among which	
83	Beyond my words can describe	
84	Despite the fact that	
85	From helping people	
86	Only in such a way can we…	
87	As a result	
88	At beginning	
89	At the very point	
90	In a…way	
91	In my opinion	
92	In my view	
93	In our times	
94	In some degree	
95	Just around the corner	
96	Out of date	
97	Under police guard	
98	with day and night	
99	Without feeling	
100	It is important to	
101	It seems	
102	It’s no doubt that	
103	It’s no use of	
104	It’s of great importance to	
105	It’s very vital for us to	
106	It’s widely acknowledged that	
107	Are reluctant to	

108	Be full of	
109	Be important for…	
110	Be of vital importance to	
111	Be the one of the	
112	Be tired of	
113	Drummed into	
114	Take place	
115	Be acknowledged	
116	Be born to	
117	Be associated with	
118	Be drawn into	
119	Be equipped with	
120	Be filled with	
121	Be originated in	
122	Born with courage	
123	We'd better	
124	Here are reasons	
125	Here is as follows	
126	It is a fine time to	
127	It is a time of	
128	It is a time to	
129	It was not once I heard about	
130	It's said	
131	The tradition has been endowed with	
132	There are a lot of	
133	There is a hot issue	
134	There is a saying	
135	There is no distinct between	
136	There is no doubt that	
137	There is no need much evidence to	
138	There is saying	
139	Account for	
140	Aim to	
141	Arise one's attention to	
142	Attach great importance	
143	Attach the greatest importance of	
144	be due to	
145	Be still going on	
146	Bear the burden of	
147	Break apart	
148	Break up with	
149	Bring good luck	
150	Burned to death	
151	Call for	

152	Call on	
153	Caused a great concern to	
154	Come across	
155	Come to realise	
156	Concentrate on	
157	Consider…as	
158	Contribute to	
159	Dated back to	
160	Deal with	
161	Destined to	
162	During the festival	
163	Enjoy in	
164	Enjoy themselves	
165	Escape from	
166	Face with	
167	Fall in love with	
168	Fall on	
169	Feel sorry for	
170	Find someone being	
171	Fly kite	
172	Gear down	
173	Go bad	
174	Go out of	
175	Had some similarities	
176	Have a crisis on	
177	Have little related to	
178	Keep…clean	
179	Live in	
180	Make it possible for	
181	Pay attention to	
182	Pay respect to	
183	People are accustomed to	
184	People are willing to go out	
185	Play an essential important role	
186	Play an important role	
187	Prefer to lead	
188	Put out	
189	Put us in danger	
190	Quarrel for	
191	Regard it as	
192	Remind us of	
193	Ride on swing	
194	Right the wrong	
195	Seek fortune	

196	Set fire for food	
197	Set fire to	
198	Set my mind to	
199	Show great concern to	
200	Show his respect to	
201	Spare our time enjoying	
202	Spring plough	
203	Start from	
204	Stay in touch	
205	Stop blaming	
206	Stumble across	
207	Succeed in	
208	Take a breath	
209	Take on a new look	
210	Take spring outing	
211	Taking photographs	
212	Tell apart	
213	Told…not to do	
214	Transfer into	
215	Transform into	
216	Treasure it as	
217	Treat…as…	
218	Used to	
219	Wear willow branches on the head	
220	Wrapped for burial	
	Moreover	Individual word
	However	Individual word
	And	Individual word
	Handcuffed	Individual word
	Ex-welder	Individual word
	Except	Individual word
	Meanwhile	Individual word
	Unforgettable	Individual word
	Grief-stricken	Individual word
	Furthermore	Individual word
	Short-time	Individual word
	Above all	Overlapped
	Account for	Overlapped
	As far as I am concerned	Overlapped
	As well as	Overlapped
	Call for	Overlapped
	Despite the fact that	Overlapped
	First and foremost	Overlapped
	In my opinion	Overlapped

	In my point of view	Overlapped
	In spite of	Overlapped
	It is no doubt that	Overlapped
	Rather than	Overlapped
	Set fire to	Overlapped
	There is no doubt that	Overlapped
	The last but not least	Subsumed
	Try our best to	Subsumed
	Be accustomed to	Subsumed
	Be willing to	Subsumed

Appendix 11 Four-word clusters extracted by corpus linguistics approach

Year 1 four-word clusters

N	cluster	Freq.	%	Texts	%
1	the dragon boat festival	217	0.39	79	26.33
2	dragon boat festival is	129	0.23	74	24.67
3	one of the most	55	0.10	52	17.33
4	is one of the	51	0.09	45	15.00
5	of the most important	47	0.08	45	15.00
6	festival is one of	45	0.08	41	13.67
7	boat festival is one	44	0.08	40	13.33
8	the most important traditional	32	0.06	32	10.67
9	I was born in	27	0.05	27	9.00
10	important traditional festivals in	27	0.05	26	8.67
11	most important traditional festivals	26	0.05	26	8.67
12	more and more people	24	0.04	22	7.33
13	my English name is	22	0.04	22	7.33
14	traditional festivals in china	20	0.04	20	6.67
15	boat festival is a	20	0.04	18	6.00
16	I want to be	19	0.03	14	4.67
17	as far as I	18	0.03	18	6.00
18	that the dragon boat	18	0.03	16	5.33
19	I went to the	18	0.03	10	3.33
20	as is known to	17	0.03	17	5.67
21	want to be a	17	0.03	14	4.67
22	try my best to	15	0.03	14	4.67
23	far as I am	14	0.02	14	4.67
24	there is no doubt	14	0.02	14	4.67
25	about the dragon boat	14	0.02	12	4.00
26	Hubei university of automotive	13	0.02	13	4.33
27	is no doubt that	13	0.02	13	4.33
28	of the dragon boat	13	0.02	11	3.67
29	university of automotive technology	12	0.02	12	4.00
30	when it comes to	12	0.02	12	4.00
31	try our best to	12	0.02	11	3.67
32	people who are in	12	0.02	9	3.00
33	and my English name	11	0.02	11	3.67
34	I think it is	11	0.02	9	3.00
35	dragon boat festival in	11	0.02	8	2.67
36	to help those in	11	0.02	8	2.67
37	when I was six	10	0.02	10	3.33
38	so I want to	10	0.02	8	2.67
39	I was six years	9	0.02	9	3.00
40	in the dragon boat	9	0.02	9	3.00
41	jumped into the river	9	0.02	9	3.00
42	it is said that	9	0.02	8	2.67
43	who are in need	9	0.02	8	2.67
44	to be a volunteer	9	0.02	7	2.33
45	help those in need	9	0.02	6	2.00
46	if you want to	9	0.02	6	2.00
47	to be a teacher	9	0.02	5	1.67
48	his country very much	8	0.01	8	2.67
49	is international volunteers day	8	0.01	8	2.67

50	is not only a	8	0.01	8	2.67
51	it is a good	8	0.01	8	2.67
52	not only can we	8	0.01	8	2.67
53	was born in a	8	0.01	8	2.67
54	when I was a	8	0.01	8	2.67
55	is the dragon boat	8	0.01	7	2.33
56	the benefits of volunteering	8	0.01	7	2.33
57	when I was in	8	0.01	7	2.33
58	I'll try my best	8	0.01	6	2.00
59	and more people are	7	0.01	7	2.33
60	at the same time	7	0.01	7	2.33
61	can learn a lot	7	0.01	7	2.33
62	dragon boat festival to	7	0.01	7	2.33
63	he jumped into the	7	0.01	7	2.33
64	I major in English	7	0.01	7	2.33
65	I went to school	7	0.01	7	2.33
66	into the Miluo river	7	0.01	7	2.33
67	m an English major	7	0.01	7	2.33
68	million registered volunteers who	7	0.01	7	2.33
69	my major is English	7	0.01	7	2.33
70	my name is Wang	7	0.01	7	2.33
71	people around the world	7	0.01	7	2.33
72	registered volunteers who have	7	0.01	7	2.33
73	the dragon boat competition	7	0.01	7	2.33
74	a volunteer is a	7	0.01	6	2.00
75	celebrate the dragon boat	7	0.01	6	2.00
76	in china it is	7	0.01	6	2.00
77	billion hours of volunteer	6	0.01	6	2.00
78	boat festival is an	6	0.01	6	2.00
79	boat festival is the	6	0.01	6	2.00
80	festival is a traditional	6	0.01	6	2.00
81	festivals in china it	6	0.01	6	2.00
82	have provided more than	6	0.01	6	2.00
83	have volunteered to help	6	0.01	6	2.00
84	hours of volunteer work	6	0.01	6	2.00
85	I come from class	6	0.01	6	2.00
86	I hope I can	6	0.01	6	2.00
87	I went to a	6	0.01	6	2.00
88	I went to primary	6	0.01	6	2.00
89	important traditional festival in	6	0.01	6	2.00
90	into the river to	6	0.01	6	2.00
91	is a good way	6	0.01	6	2.00
92	is a traditional festival	6	0.01	6	2.00
93	is known to us	6	0.01	6	2.00
94	it is not only	6	0.01	6	2.00
95	jumped into the Miluo	6	0.01	6	2.00
96	loved his country very	6	0.01	6	2.00
97	millions of people around	6	0.01	6	2.00
98	most important festival in	6	0.01	6	2.00
99	my dream is to	6	0.01	6	2.00
100	observe the festival in	6	0.01	6	2.00
101	of millions of people	6	0.01	6	2.00
102	of people around the	6	0.01	6	2.00
103	people observe the festival	6	0.01	6	2.00

104	should try our best	6	0.01	6	2.00
105	tens of millions of	6	0.01	6	2.00
106	the first time i	6	0.01	6	2.00
107	the most important festival	6	0.01	6	2.00
108	the most important festivals	6	0.01	6	2.00
109	those who are in	6	0.01	6	2.00
110	to go to school	6	0.01	6	2.00
111	to help those who	6	0.01	6	2.00
112	to take part in	6	0.01	6	2.00
113	to the primary school	6	0.01	6	2.00
114	traditional festival in china	6	0.01	6	2.00
115	traditional festivals in our	6	0.01	6	2.00
116	volunteers who have provided	6	0.01	6	2.00
117	you can call me	6	0.01	6	2.00
118	being a volunteer is	6	0.01	5	1.67
119	to observe this festival	6	0.01	5	1.67
120	when the dragon boat	6	0.01	5	1.67
121	a lot of things	5		5	1.67
122	a traditional festival in	5		5	1.67
123	all over the world	5		5	1.67
124	around the world have	5		5	1.67
125	become more and more	5		5	1.67
126	boat festival is not	5		5	1.67
127	festival is a very	5		5	1.67
128	festival is not only	5		5	1.67
129	festivals in our country	5		5	1.67
130	help those who are	5		5	1.67
131	I was five years	5		5	1.67
132	I was sent to	5		5	1.67
133	in memory of the	5		5	1.67
134	is in memory of	5		5	1.67
135	is my English name	5		5	1.67
136	learn a lot of	5		5	1.67
137	like to make friends	5		5	1.67
138	more and more popular	5		5	1.67
139	most important traditional festival	5		5	1.67
140	necessary for us to	5		5	1.67
141	on the dragon boat	5		5	1.67
142	our best to help	5		5	1.67
143	school when I was	5		5	1.67
144	should pay more attention	5		5	1.67
145	the hometown of Qu	5		5	1.67
146	the world have volunteered	5		5	1.67
147	to make friends with	5		5	1.67
148	to observe the festival	5		5	1.67
149	to the middle school	5		5	1.67
150	volunteered to help those	5		5	1.67
151	we should pay more	5		5	1.67
152	we should try our	5		5	1.67
153	went to primary school	5		5	1.67
154	when I was five	5		5	1.67
155	which is said to	5		5	1.67
156	who have provided more	5		5	1.67
157	with the development of	5		5	1.67

158	world have volunteered to	5		5	1.67

Year 3 four-word clusters

N	Word	Freq.	%	Texts	%
1	the Qing Ming festival	67	0.10	25	9.96
2	Qing Ming festival is	48	0.07	27	10.76
3	we should help strangers	39	0.06	31	12.35
4	more and more people	28	0.04	26	10.36
5	that we should help	24	0.04	20	7.97
6	should we help strangers	23	0.04	21	8.37
7	is one of the	21	0.03	19	7.57
8	as far as I	20	0.03	20	7.97
9	with the development of	19	0.03	19	7.57
10	willing to read classics	19	0.03	16	6.37
11	far as I am	17	0.03	17	6.77
12	the importance of reading	17	0.03	15	5.98
13	people are willing to	16	0.02	16	6.37
14	is very important for	15	0.02	12	4.78
15	are willing to read	14	0.02	14	5.58
16	festival is one of	14	0.02	14	5.58
17	try our best to	14	0.02	13	5.18
18	importance of reading classics	14	0.02	12	4.78
19	reading classics is very	14	0.02	12	4.78
20	classics is very important	13	0.02	11	4.38
21	Ming festival is a	13	0.02	11	4.38
22	tomb sweeping day is	13	0.02	10	3.98
23	less and less people	12	0.02	12	4.78
24	Ming festival is one	12	0.02	12	4.78
25	our best to help	12	0.02	11	4.38
26	and more people are	11	0.02	11	4.38
27	is not only a	11	0.02	11	4.38
28	is the most important	11	0.02	11	4.38
29	Qing Ming is a	11	0.02	10	3.98
30	I think we should	11	0.02	9	3.59
31	we should try our	10	0.02	10	3.98
32	is a time to	10	0.02	9	3.59
33	for us to read	10	0.02	6	2.39
34	a combination of sadness	9	0.01	9	3.59
35	as is known to	9	0.01	9	3.59
36	combination of sadness and	9	0.01	9	3.59
37	seasonal division points in	9	0.01	9	3.59
38	should try our best	9	0.01	9	3.59
39	the temperature will rise	9	0.01	9	3.59
40	give a hand to	9	0.01	8	3.19
41	is no doubt that	9	0.01	8	3.19
42	a period to honour	8	0.01	8	3.19
43	ancestors and sweep the	8	0.01	8	3.19
44	bow before the memorial	8	0.01	8	3.19
45	each year after the	8	0.01	8	3.19
46	helping others is a	8	0.01	8	3.19
47	is the high time	8	0.01	8	3.19
48	it is the high	8	0.01	8	3.19
49	learn to protect ourselves	8	0.01	8	3.19

50	less people are willing	8	0.01	8	3.19
51	of sadness and happiness	8	0.01	8	3.19
52	one of the most	8	0.01	8	3.19
53	the high time for	8	0.01	8	3.19
54	to read classics as	8	0.01	8	3.19
55	when it comes to	8	0.01	8	3.19
56	take advantage of our	8	0.01	7	2.79
57	there is no doubt	8	0.01	7	2.79
58	a festival of commemoration	7	0.01	7	2.79
59	a time to remember	7	0.01	7	2.79
60	and less people are	7	0.01	7	2.79
61	and sweep the tombs	7	0.01	7	2.79
62	contrast to the sadness	7	0.01	7	2.79
63	extended the day to	7	0.01	7	2.79
64	festival sees a combination	7	0.01	7	2.79
65	in contrast to the	7	0.01	7	2.79
66	in my point of	7	0.01	7	2.79
67	is a period to	7	0.01	7	2.79
68	make our society more	7	0.01	7	2.79
69	Ming festival is also	7	0.01	7	2.79
70	Ming festival sees a	7	0.01	7	2.79
71	no doubt that we	7	0.01	7	2.79
72	period to honour and	7	0.01	7	2.79
73	Qing Ming festival as	7	0.01	7	2.79
74	Qing Ming festival sees	7	0.01	7	2.79
75	rise up and rainfall	7	0.01	7	2.79
76	sadness of the tomb	7	0.01	7	2.79
77	sweep the tombs of	7	0.01	7	2.79
78	temperature will rise up	7	0.01	7	2.79
79	the development of the	7	0.01	7	2.79
80	the sadness of the	7	0.01	7	2.79
81	their ancestors and sweep	7	0.01	7	2.79
82	this is the most	7	0.01	7	2.79
83	time to read classics	7	0.01	7	2.79
84	time to remember the	7	0.01	7	2.79
85	to read classics the	7	0.01	7	2.79
86	to remember the dead	7	0.01	7	2.79
87	to the sadness of	7	0.01	7	2.79
88	to their ancestors and	7	0.01	7	2.79
89	will rise up and	7	0.01	7	2.79
90	festival is a time	7	0.01	6	2.39
91	of food and spirit	7	0.01	6	2.39
92	very important for us	7	0.01	6	2.39
93	a seasonal point to	6		6	2.39
94	a time to plant	6		6	2.39
95	advantage of our sympathy	6		6	2.39
96	ancestors often extended the	6		6	2.39
97	and bow before the	6		6	2.39
98	and only cold food	6		6	2.39
99	and paper money and	6		6	2.39
100	and rainfall increases it	6		6	2.39
101	as our ancestors often	6		6	2.39
102	at the same time	6		6	2.39
103	be very cautious of	6		6	2.39

104	before the memorial tablet	6	6	2.39
105	burn incense and paper	6	6	2.39
106	but the Qing Ming	6	6	2.39
107	classics as a college	6	6	2.39
108	cold food is served	6	6	2.39
109	commemoration the Qing Ming	6	6	2.39
110	cook on this day	6	6	2.39
111	day and only cold	6	6	2.39
112	festival as our ancestors	6	6	2.39
113	festival is not only	6	6	2.39
114	festival of commemoration the	6	6	2.39
115	have little time to	6	6	2.39
116	helping others can sometimes	6	6	2.39
117	helping others is helping	6	6	2.39
118	helping strangers is a	6	6	2.39
119	high time for spring	6	6	2.39
120	hope of spring on	6	6	2.39
121	important day of sacrifice	6	6	2.39
122	incense and paper money	6	6	2.39
123	increases it is the	6	6	2.39
124	is more a festival	6	6	2.39
125	it is said that	6	6	2.39
126	Ming festival is not	6	6	2.39
127	money and bow before	6	6	2.39
128	more a festival of	6	6	2.39
129	most important day of	6	6	2.39
130	not cook on this	6	6	2.39
131	not only a seasonal	6	6	2.39
132	of commemoration the Qing	6	6	2.39
133	of spring on this	6	6	2.39
134	of the most important	6	6	2.39
135	offer sacrifices to their	6	6	2.39
136	often extended the day	6	6	2.39
137	on the way to	6	6	2.39
138	on this day and	6	6	2.39
139	only a seasonal point	6	6	2.39
140	only cold food is	6	6	2.39
141	our ancestors often extended	6	6	2.39
142	our society more harmonious	6	6	2.39
143	paper money and bow	6	6	2.39
144	point to guide farm	6	6	2.39
145	Qing Ming festival was	6	6	2.39
146	rainfall increases it is	6	6	2.39
147	reading classics we can	6	6	2.39
148	remember the dead and	6	6	2.39
149	sacrifices to their ancestors	6	6	2.39
150	seasonal point to guide	6	6	2.39
151	sees a combination of	6	6	2.39
152	so that we can	6	6	2.39
153	sowing but the Qing	6	6	2.39
154	spirits of deceased ancestors	6	6	2.39
155	spring on this day	6	6	2.39
156	tend to be very	6	6	2.39
157	that helping others can	6	6	2.39

158	the beginning of the	6	6	2.39
159	the dead and the	6	6	2.39
160	the most important day	6	6	2.39
161	the spirits of deceased	6	6	2.39
162	the tombs of the	6	6	2.39
163	they were later combined	6	6	2.39
164	they will not cook	6	6	2.39
165	this day and only	6	6	2.39
166	to be very cautious	6	6	2.39
167	to give a hand	6	6	2.39
168	to help strangers in	6	6	2.39
169	we should help others	6	6	2.39
170	we should try to	6	6	2.39
171	when the lord succeeded	6	6	2.39
172	when you are in	6	6	2.39
173	whether we should help	6	6	2.39
174	will not cook on	6	6	2.39
175	fewer and fewer people	6	5	1.99
176	food and spirit money	6	5	1.99
177	a piece of his	5	5	1.99
178	also a time to	5	5	1.99
179	also learn to protect	5	5	1.99
180	and favourites of the	5	5	1.99
181	and happiness this is	5	5	1.99
182	and minority ethnic groups	5	5	1.99
183	and pay respect to	5	5	1.99
184	and sowing but the	5	5	1.99
185	and trees grow fast	5	5	1.99
186	at this time offer	5	5	1.99
187	become more and more	5	5	1.99
188	becoming more and more	5	5	1.99
189	becoming the ruler of	5	5	1.99
190	before the Qing Ming	5	5	1.99
191	both the Han and	5	5	1.99
192	by serving a piece	5	5	1.99
193	day after the winter	5	5	1.99
194	day before the Qing	5	5	1.99
195	doubt that we should	5	5	1.99
196	festival was called arbour	5	5	1.99
197	festival was usually one	5	5	1.99
198	flowers and favourites of	5	5	1.99
199	from my point of	5	5	1.99
200	go to the library	5	5	1.99
201	growing number of people	5	5	1.99
202	Han and minority ethnic	5	5	1.99
203	happiness this is the	5	5	1.99
204	has always been considered	5	5	1.99
205	have a lot of	5	5	1.99
206	high and trees grow	5	5	1.99
207	important for us to	5	5	1.99
208	in becoming the ruler	5	5	1.99
209	in our daily life	5	5	1.99
210	in Shanxi province in	5	5	1.99
211	is high and trees	5	5	1.99

212	is very important to	5	5	1.99
213	it is a period	5	5	1.99
214	it is a time	5	5	1.99
215	it is more a	5	5	1.99
216	know the importance of	5	5	1.99
217	leg when the lord	5	5	1.99
218	lived in Shanxi province	5	5	1.99
219	lord ordered all fires	5	5	1.99
220	lord succeeded in becoming	5	5	1.99
221	love to fly kites	5	5	1.99
222	maybe the best of	5	5	1.99
223	Ming festival as our	5	5	1.99
224	Ming festival was called	5	5	1.99
225	Ming is a time	5	5	1.99
226	minority ethnic groups at	5	5	1.99
227	money could keep them	5	5	1.99
228	no choice but to	5	5	1.99
229	on the anniversary of	5	5	1.99
230	one day before the	5	5	1.99
231	ordered all fires in	5	5	1.99
232	out on the anniversary	5	5	1.99
233	pay more attention to	5	5	1.99
234	people are afraid of	5	5	1.99
235	piece of his own	5	5	1.99
236	protect ourselves from being	5	5	1.99
237	remain where he was	5	5	1.99
238	s life by serving	5	5	1.99
239	s no doubt that	5	5	1.99
240	sadness and happiness this	5	5	1.99
241	serving a piece of	5	5	1.99
242	should also learn to	5	5	1.99
243	succeeded in becoming the	5	5	1.99
244	sweep tombs and offer	5	5	1.99
245	that the spirits of	5	5	1.99
246	the custom of spring	5	5	1.99
247	the Han and minority	5	5	1.99
248	the high pace of	5	5	1.99
249	the lord ordered all	5	5	1.99
250	the lord succeeded in	5	5	1.99
251	the memorial tablet in	5	5	1.99
252	the survival rate of	5	5	1.99
253	the tomb sweeping day	5	5	1.99
254	the trees and grass	5	5	1.99
255	think reading classics is	5	5	1.99
256	time offer sacrifices to	5	5	1.99
257	to death to commemorate	5	5	1.99
258	to fly kites during	5	5	1.99
259	to pay respect to	5	5	1.99
260	to protect ourselves from	5	5	1.99
261	to read classics and	5	5	1.99
262	to read the classics	5	5	1.99
263	to remain where he	5	5	1.99
264	to sweep tombs and	5	5	1.99
265	tombs and offer sacrifices	5	5	1.99

266	tombs of the diseased	5	5	1.99
267	up and rainfall increases	5	5	1.99
268	usually one day before	5	5	1.99
269	was usually one day	5	5	1.99
270	we should also learn	5	5	1.99
271	where he was and	5	5	1.99
272	who are in trouble	5	5	1.99
273	who are willing to	5	5	1.99
274	who lived in Shanxi	5	5	1.99
275	who take advantage of	5	5	1.99
276	you are in trouble	5	5	1.99

Appendix 12 Structural and Functional categorisation of formulaic strings

Year 1 Structural and Functional categorisation of formulaic strings

No.	strings	structure	sub-structure	function	sub-function
1	As well	other	ill-formed	textual	transition
2	…be drawn	vp	passive verb	ideational	action
3	a freshman majoring in	np	np+of	ideational	description
4	A kind of	np	np+of	ideational	quantification
5	a money mighty world	np	other np	ideational	description
6	A strong feeling of being	np	np+of	ideational	description
7	A three-day holiday	np	other NP	ideational	description
8	A very place where hold	np	np+ other post modifier	ideational	location
9	According to	pp	other PP	textual	resultative
10	achieve my dreams	vp	active VP	ideational	action
11	After all	pp	other PP	textual	transition
12	After graduation	pp	other PP	ideational	description
13	After that	pp	other PP	ideational	location
14	All kinds of	np	np+of	ideational	quantification
15	And so forth	other	ill-formed	textual	exemplification
16	Are fulfilled	VP	passive verb	ideational	description
17	As a Chinese	PP	other PP	ideational	description
18	As a result	PP	other PP	textual	resultative
19	As an old saying goes	clausal	as-clause	textual	generalisation
20	As before	PP	other PP	textual	transition
21	As far as I am concerned	pp	other PP	textual	framing
22	As for	pp	other PP	textual	framing
23	As is known to all	clausal	as-clause	textual	generalisation
24	As it is known to us all	clausal	as-clause	textual	generalisation
25	As it is said	clausal	as-clause	textual	generalisation
26	As their tradition	pp	other PP	ideational	description
27	As to	pp	other PP	textual	framing
28	As we all know	clausal	as-clause	textual	generalisation
29	At that time	pp	other PP	ideational	location
30	At the same time	pp	other PP	textual	transition
31	At the side of	pp	pp+of	ideational	location
32	Baby doves	np	other NP	ideational	description
33	Be admitted as	vp	passive verb+pp	ideational	description
34	Be faced with	vp	passive verb+pp	ideational	description
35	Be filled up with	vp	passive verb+pp	ideational	description
36	Be given	vp	passive verb	ideational	description

37	Be responsible for	vp	copula be+np/adj.p	interpersonal	stance
38	be so honour to	clausal	to-clause	interpersonal	stance
39	Because of	pp	other PP	textual	resultative
40	Been asking for perfect	vp	active VP	ideational	action
41	Benefit from it	vp	active VP	ideational	action
42	Best occupation	np	other NP	ideational	description
43	Bring…up	vp	active VP	ideational	action
44	Build up my body	vp	active VP	ideational	action
45	By the way	pp	other PP	textual	transition
46	Come to our mind	vp	active VP	ideational	action
47	Comparing with	vp	active VP	textual	transition
48	Consist of	vp	active VP	ideational	action
49	Devoted himself	vp	active VP	ideational	action
50	did well in	vp	active VP	ideational	action
51	Die for	vp	active VP	ideational	action
52	Drive evil and ghost away	vp	active VP	ideational	action
53	Earn money to support	vp	active VP	ideational	action
54	Even though	adj.p/adv.p	adv.p	textual	transition
55	Family quality time	np	other NP	ideational	description
56	Finish it well	vp	active VP	ideational	action
57	First…second…last but not least	other	template	textual	structuring
58	Firstly...Secondly…Finally	other	template	textual	structuring
59	For example	pp	other PP	textual	exemplification
60	For instance	pp	other PP	textual	exemplification
61	For one thing…for another	other	template	textual	transition
62	From my point of view	pp	pp+of	textual	framing
63	from then on	pp	other PP	ideational	location
64	Gain something good from it	vp	active VP	ideational	action
65	get along very well with	vp	active VP	ideational	action
66	Get home	vp	active VP	ideational	action
67	Get something useful	vp	active VP	ideational	action
68	Get together	vp	active VP	ideational	action
69	Get up very early	vp	active VP	ideational	action
70	Give me a deep impression on	vp	active VP	ideational	action
71	Give their hand to	vp	active VP	ideational	action

72	Go away	vp	active VP	ideational	action
73	Go back to	vp	active VP	ideational	action
74	Go to the river	vp	active VP	ideational	action
75	Had no time	vp	active VP	ideational	action
76	Have trouble communicating	vp	active VP	ideational	action
77	Having delivered myself about	vp	active VP	textual	framing
78	High ranks	np	other NP	ideational	description
79	How beautiful the scenery is	clausal	wh-clause	interpersonal	stance
80	How exciting	clausal	wh-clause	interpersonal	stance
81	How exciting game it is	clausal	wh-clause	interpersonal	stance
82	Hurried to	vp	active VP	ideational	action
83	I am behalf of…	vp	pronoun/np+be …	interpersonal	stance
84	I am good at	vp	pronoun/np+be …	interpersonal	stance
85	I believe that	clausal	1st/2nd person pronoun+dep. cl.	interpersonal	stance
86	I hold the view that	clausal	1st/2nd person pronoun+dep.cl .	interpersonal	stance
87	I hope I'll	clausal	1st/2nd person pronoun+dep.cl .	interpersonal	stance
88	I think it's a	clausal	1st/2nd person pronoun+dep.cl .	interpersonal	stance
89	I was admitted to	vp	pronoun/np+be …	ideational	description
90	I'll try my best to be impressive	vp	pronoun/np+be …	interpersonal	stance
91	I'm not skilled at	vp	pronoun/np+be …	interpersonal	stance
92	If permitted	clausal	if-clause	textual	resultative
93	If you want to be loved, love others.	clausal	if-clause	interpersonal	engagement
94	If you…you will	clausal	if-clause	interpersonal	engagement
95	If...I'll be	clausal	if-clause	interpersonal	stance
96	Improve your skills	vp	active VP	interpersonal	engagement
97	In a word	pp	other PP	textual	structuring
98	In addition	pp	other PP	textual	transition
99	In conclusion	pp	other PP	textual	structuring
100	In fact	pp	other PP	textual	transition
101	in honour of	pp	other PP	ideational	description
102	In lunar	pp	other PP	ideational	location

103	In order to	pp	other PP	textual	resultative
104	In other words	pp	other PP	textual	transition
105	In recent years	pp	other PP	ideational	location
106	In the ancient China	pp	other PP	ideational	location
107	In the meantime	pp	other PP	ideational	location
108	In the morning	pp	other PP	ideational	location
109	In the river	pp	other PP	ideational	location
110	Is far away from…	vp	copula be+np/adj.p	ideational	description
111	It is a special day for us to	vp	pronoun/np+be …	interpersonal	stance
112	it is my pleasure to be	vp	anticipatory it	interpersonal	stance
113	It is necessary for	vp	anticipatory it	interpersonal	stance
114	It was time for me	vp	pronoun/np+be …	interpersonal	stance
115	It would be much happier of giving instead of receiving	vp	anticipatory it	interpersonal	stance
116	It's said that you can judge a man from his hobbits	vp	pronoun/np+be …	interpersonal	engagement
117	Join in	vp	active VP	ideational	action
118	Jump into	vp	active VP	ideational	action
119	Keep…from	vp	active VP	ideational	action
120	Know what happened around	clausal	wh-clause	ideational	action
121	learn a lot from	vp	active VP	ideational	action
122	Learn…well	vp	active VP	ideational	action
123	Less and less	vp	active VP	textual	transition
124	Look upon	vp	active VP	ideational	action
125	Lose weight	vp	active VP	ideational	action
126	Made great contribution to	vp	active VP	ideational	action
127	Make my dream come true	vp	active VP	ideational	action
128	Make new friends with	vp	active VP	ideational	action
129	Make…into	vp	active VP	ideational	action
130	Maybe you know	other	other	interpersonal	engagement
131	More than	adj.p/adv.p	adv.p	textual	transition
132	Most of	adj.p/adv.p	adv.p	ideational	quantification
133	Much more… than…	adj.p/adv.p	adv.p	textual	transition
134	Nice to meet you	other	formulae	conversational	politeness
135	No matter how	clausal	wh-clause	interpersonal	stance

136	no pain, no gain	other	formulae	interpersonal	stance
137	Not only…but also	other	template	textual	transition
138	Now that	adj.p/adv.p	adv.p	textual	resultative
139	Of course	pp	other PP	interpersonal	stance
140	On earth	pp	other PP	interpersonal	stance
141	On the one hand…on the other hand	other	template	textual	transition
142	On the other hand	pp	other PP	textual	transition
143	One of the	np	np+of	ideational	quantification
144	one's loyalty to	vp	active VP	ideational	description
145	Open my eyes	vp	active VP	ideational	action
146	Out of kindness	pp	pp+of	ideational	description
147	Out of trouble	pp	pp+of	ideational	description
148	Pay attention to	vp	active VP	ideational	action
149	Prefer to	vp	active VP	ideational	action
150	Primary school	np	other NP	ideational	description
151	Protect…from	vp	active VP	ideational	action
152	Put it differently	vp	active VP	textual	framing
153	Put up	vp	active VP	ideational	action
154	Put…into	vp	active VP	ideational	action
155	Put…into practice	vp	active VP	ideational	action
156	Realise dreams	vp	active VP	ideational	action
157	Sacrifice…for	vp	active VP	ideational	action
158	Six years later	np	other NP	ideational	location
159	So great to the ear	clausal	to-clause	interpersonal	stance
160	So…that…	other	template	textual	resultative
161	Some of	adj.p/adv.p	adv.p	ideational	quantification
162	Something about it	np	np+ other post modifier	ideational	description
163	Something helpful	np	np+ other post modifier	ideational	description
164	Speaking the…	vp	active VP	textual	framing
165	Such as	vp	active VP	textual	exemplification
166	Support my family	vp	active VP	ideational	action
167	Surprise may be,	vp	active VP	textual	resultative
168	Take a trip in honour of	vp	active VP	ideational	action
169	Take care of	vp	active VP	ideational	action
170	Take part in	vp	active VP	ideational	action
171	Talk about	vp	active VP	ideational	action
172	Talk with	vp	active VP	ideational	action
173	Taste sweet in heart	vp	active VP	ideational	action
174	Tend to be	vp	active VP	textual	resultative

175	That would be	vp	pronoun/np+be …	textual	resultative
176	that's mean	vp	pronoun/np+be …	textual	transition
177	That's to say	VP	pronoun/np+be …	textual	resultative
178	The air is also fresh for us to breathe	other	other	interpersonal	stance
179	The bones of fish	np	np+of	ideational	description
180	The delight of giving	np	np+of	ideational	description
181	The Dragon boat festival	np	other np	ideational	description
182	The food is to Chinese people what the turkey to the Americans	other	formulae (analogy)	interpersonal	stance
183	There are	vp	pronoun/np+be …	ideational	quantification
184	There might be	vp	pronoun/np+be …	textual	resultative
185	Think of	vp	active VP	ideational	action
186	This is to say	vp	pronoun/np+be …	textual	transition
187	Those who	np	np+ other post modifier	ideational	description
188	those who are needed help	np	np+ other postmodifier	ideational	description
189	To affect everyone	clausal	to-clause	ideational	action
190	To be a…is	clausal	to-clause	interpersonal	stance
191	To begin with	vp	active VP	textual	structuring
192	To memory this great person	clausal	to-clause	ideational	action
193	Traditional festivals	np	other NP	ideational	description
194	Try our best	vp	active VP	ideational	action
195	universal love	np	other NP	ideational	description
196	Used to be	vp	active VP	ideational	description
197	Very much	adj.p/adv.p	adv.p	ideational	quantification
198	Want to honour	vp	active VP	ideational	action
199	What a bad luck!	clausal	wh-clause	interpersonal	stance
200	What a busy day	clausal	wh-clause	interpersonal	stance
201	What a lovely day	clausal	wh-clause	interpersonal	stance
202	What about my dreams?	clausal	wh-clause	conversational	simple inquiry
203	What you have learn	clausal	wh-clause	interpersonal	engagement
204	What's more	other	ill-formed	textual	transition
205	When it comes to	clausal	wh-clause	textual	framing

206	Where there is river, there is a city	other	formulae	interpersonal	stance
207	Which was said to	clausal	wh-clause	textual	transition
208	With time going by	pp	other PP	ideational	description
209	You are due to	vp	pronoun/np+be …	interpersonal	engagement
210	you can call me	vp	active VP	interpersonal	engagement

Year 3 Structural and Functional categorisation of formulaic strings

No.	**strings**	**structure**	**sub-structure**	**function**	**subfunction**
1	More important	adj.p/adv.p	adj.p	textual	transition
2	Not familiar with	adj.p/adv.p	adj.p	ideational	description
3	First and foremost	adj.p/adv.p	adv.p	textual	structuring
4	Not long ago	adj.p/adv.p	adv.p	ideational	location
5	Getting fewer and fewer	adj.p/adv.p	adv.p	textual	transition
6	First of all	adj.p/adv.p	adv.p	textual	structuring
7	Last but not least	adj.p/adv.p	adv.p	textual	structuring
8	Later on	adj.p/adv.p	adv.p	ideational	location
9	More and more	adj.p/adv.p	adv.p	textual	transition
10	More or less	adj.p/adv.p	adv.p	ideational	quantification
11	Rather than	adj.p/adv.p	adv.p	textual	transition
12	Sooner or later	adj.p/adv.p	adv.p	ideational	location
13	The last but not least	adj.p/adv.p	adv.p	textual	structuring
14	Not all	adj.p/adv.p	adv.p	ideational	quantification
15	I firmly believe that we	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
16	One never knows that	clausal	1st/2nd person pronoun+dep.cl.	textual	framing
17	We must be clear that	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
18	As it is known	clausal	as-clause	textual	generalisation
19	We, as college students, should be	clausal	as-clause	interpersonal	stance

20	Not only…but also…	other	template	textual	transition
21	Only when…can we	other	template	textual	resultative
22	Lucky enough to	clausal	to-clause	ideational	description
23	Try one's best to	Phrasal	VP	ideational	action
24	See what we are doing	clausal	wh-clause	ideational	action
25	what attracted each other maybe	clausal	wh-clause	ideational	description
26	What impressed me most	clausal	wh-clause	interpersonal	stance
27	What you should do is to	clausal	wh-clause	interpersonal	engagement
28	which is	clausal	wh-clause	textual	transition
29	which strongly illustrated…	clausal	wh-clause	textual	exemplification
30	Pure Brightness festival	np	other NP	ideational	description
31	Qing Ming Festival	np	other NP	ideational	description
32	Tomb sweeping day	np	other NP	ideational	description
33	Dreams of becoming	np	np+of	ideational	description
34	Lack of	np	np+of	ideational	quantification
35	One of the most important	np	np+of	ideational	quantification
36	Return of colour	np	np+of	ideational	description
37	The passing from one generation to another	np	np+other post modifier	ideational	description
38	Those in need	np	np+other post modifier	ideational	description
39	A three days off	np	other NP	ideational	description
40	Ancestral tablets	np	other NP	ideational	description
41	Heartless and soulless	np	other NP	ideational	description
42	Jin Wen Gong	np	other NP	ideational	description
43	The coming of an era of	np	other NP	ideational	description
44	Unexpected twists and switchbacks	np	other NP	ideational	description
45	Various choices	np	other NP	ideational	description
46	Waste-time and waste-energy thing.	np	other NP	ideational	description
47	As far as I am concerned	pp	other PP	textual	framing
48	Head over heels	other	ill-formed	ideational	description
49	To help or not to help	other	formuae	ideational	action

50	Turn a blind eye to	vp	active VP	ideational	action
51	Turn a deaf ear to	vp	active VP	ideational	action
52	What's more	other	ill-formed	textual	transition
53	No pain, no gain	other	formulae	interpersonal	stance
54	Do onto other as we would have them do onto us	other	formulae	interpersonal	stance
55	Help others is help ourselves	other	formulae	interpersonal	stance
56	Greed is the eternal soil of violence	other	other	interpersonal	stance
57	Like two stars, crashed together	other	other	ideational	description
58	Only the deaf appreciate rearing, only the blind realise the manifold blessing that life in sight	other	formulae	interpersonal	stance
59	Rose to lover, smell in hands	other	formulae	interpersonal	stance
60	The roses in her hands, the flavour in mine.	other	formulae	interpersonal	stance
61	Where there is a book, there is gold	other	formulae	interpersonal	stance
62	As well as	other	ill-formed	textual	transition
63	Few and far between	other	ill-formed	ideational	quantification
64	First…Second…Finally	other	template	textual	structuring
65	First…Then…	other	template	textual	structuring
66	To begin with…then…at last	other	template	textual	structuring
67	Hot issues vary from day to day	other	other	textual	generalisation
68	Above all	pp	other PP	textual	transition
69	According to legends	pp	other PP	textual	resultative
70	All in all	pp	other PP	textual	structuring
71	For example	pp	other PP	textual	exemplification
72	In order to	pp	other PP	textual	resultative
73	In spite of	pp	PP with embedded of phrase	textual	transition
74	In terms of	pp	PP with embedded of phrase	textual	framing

75	Instead of	pp	PP with embedded of phrase	textual	transition
76	At last	pp	other PP	textual	structuring
77	At the same time	pp	other PP	textual	transition
78	On the other hand	pp	other PP	textual	transition
79	For thousands of hundred years	pp	pp+of	ideational	quantificatio n
80	From my point of view	pp	pp+of	textual	framing
81	In my point of view	pp	pp+of	textual	framing
82	Among which	pp	pp+other post modifier	textual	transition
83	Beyond my words can describe	pp	pp+other post modifier	ideational	description
84	Despite the fact that	pp	pp+other post modifier	textual	transition
85	From helping people	pp	pp+other post modifier	ideational	action
86	Only in such a way can we…	pp	pp+other post modifier	textual	resultative
87	As a result	pp	other PP	textual	resultative
88	At beginning	pp	other PP	textual	structuring
89	At the very point	pp	other PP	ideational	location
90	In a…way	pp	other PP	ideational	description
91	In my opinion	pp	other PP	textual	framing
92	In my view	pp	other PP	textual	framing
93	In our times	pp	other PP	ideational	location
94	In some degree	pp	other PP	ideational	quantificatio n
95	Just around the corner	pp	other PP	ideational	location
96	Out of date	pp	other PP	ideational	description
97	Under police guard	pp	other PP	ideational	description
98	with day and night	pp	other PP	ideational	description
99	Without feeling	pp	other PP	ideational	description
100	It is important to	vp	anticipatory it	interpersona l	stance
101	It seems	vp	anticipatory it	interpersona l	stance
102	It’s no doubt that	vp	anticipatory it	interpersona l	stance
103	It’s no use of	vp	anticipatory it	interpersona l	stance
104	It’s of great importance to	vp	anticipatory it	interpersona l	stance
105	It’s very vital for us to	vp	anticipatory it	interpersona l	stance

106	It's widely acknowledged that	vp	anticipatory it	textual	generalisation
107	Are reluctant to	vp	copula be+ np/adj.p	ideational	description
108	Be full of	vp	copula be+ np/adj.p	ideational	description
109	Be important for…	vp	copula be+ np/adj.p	interpersonal	stance
110	Be of vital importance to	vp	copula be+ np/adj.p	interpersonal	stance
111	Be the one of the	vp	copula be+ np/adj.p	ideational	quantification
112	Be tired of	vp	copula be+ np/adj.p	ideational	description
113	Drummed into	vp	copula be+ np/adj.p	ideational	action
114	Take place	vp	active VP	ideational	action
115	Be acknowledged	vp	passive	ideational	description
116	Be born to	vp	passive	ideational	description
117	Be associated with	vp	passive+pp	textual	resultative
118	Be drawn into	vp	passive+pp	ideational	description
119	Be equipped with	vp	passive+pp	ideational	description
120	Be filled with	vp	passive+pp	ideational	description
121	Be originated in	vp	passive+pp	textual	resultative
122	Born with courage	vp	passive+pp	ideational	description
123	We'd better	vp	active VP	interpersonal	stance
124	Here are reasons	vp	pronoun/np+be…	textual	framing
125	Here is as follows	vp	pronoun/np+be…	textual	framing
126	It is a fine time to	vp	pronoun/np+be…	interpersonal	stance
127	It is a time of	vp	pronoun/np+be…	interpersonal	stance
128	It is a time to	vp	pronoun/np+be…	interpersonal	stance
129	It was not once I heard about	vp	pronoun/np+be…	textual	framing
130	It's said	vp	pronoun/np+be…	textual	framing
131	The tradition has been endowed with	vp	pronoun/np+be…	ideational	description
132	There are a lot of	vp	pronoun/np+be…	ideational	quantification
133	There is a hot issue	vp	pronoun/np+be…	textual	framing
134	There is a saying	vp	pronoun/np+be…	textual	framing
135	There is no distinct between	vp	pronoun/np+be…	textual	transition
136	There is no doubt that	vp	pronoun/np+be…	interpersonal	stance

137	There is no need much evidence to	vp	pronoun/np+be…	interpersonal	stance
138	There is saying	vp	pronoun/np+be…	textual	framing
139	Account for	vp	active VP	textual	resultative
140	Aim to	vp	active VP	ideational	action
141	Arise one's attention to	vp	active VP	ideational	action
142	Attach great importance	vp	active VP	ideational	action
143	Attach the greatest importance of	vp	active VP	ideational	action
144	be due to	vp	copula be+ np/adj.p	ideational	description
145	Be still going on	vp	copula be+ np/adj.p	ideational	description
146	Bear the burden of	vp	active VP	ideational	action
147	Break apart	vp	active VP	ideational	action
148	Break up with	vp	active VP	ideational	action
149	Bring good luck	vp	active VP	ideational	action
150	Burned to death	vp	active VP	ideational	action
151	Call for	vp	active VP	ideational	action
152	Call on	vp	active VP	ideational	action
153	Caused a great concern to	vp	active VP	ideational	action
154	Come across	vp	active VP	ideational	action
155	Come to realise	vp	active VP	ideational	action
156	Concentrate on	vp	active VP	ideational	action
157	Consider…as	vp	active VP	ideational	action
158	Contribute to	vp	active VP	ideational	action
159	Dated back to	vp	active VP	ideational	description
160	Deal with	vp	active VP	ideational	action
161	Destined to	vp	active VP	ideational	action
162	During the festival	vp	active VP	ideational	location
163	Enjoy in	vp	active VP	ideational	action
164	Enjoy themselves	vp	active VP	ideational	action
165	Escape from	vp	active VP	ideational	action
166	Face with	vp	active VP	ideational	action
167	Fall in love with	vp	active VP	ideational	action
168	Fall on	vp	active VP	ideational	action
169	Feel sorry for	vp	active VP	interpersonal	stance
170	Find someone being	vp	active VP	ideational	description
171	Fly kite	vp	active VP	ideational	action
172	Gear down	vp	active VP	ideational	action
173	Go bad	vp	active VP	ideational	description
174	Go out of	vp	active VP	ideational	action

175	Had some similarities	vp	active VP	textual	transition
176	Have a crisis on	vp	active VP	ideational	description
177	Have little related to	vp	active VP	textual	resultative
178	Keep…clean	vp	active VP	ideational	description
179	Live in	vp	active VP	ideational	action
180	Make it possible for	vp	active VP	ideational	action
181	Pay attention to	vp	active VP	ideational	action
182	Pay respect to	vp	active VP	ideational	action
183	People are accustomed to	vp	pronoun/np+be…	ideational	description
184	People are willing to go out	vp	pronoun/np+be…	ideational	description
185	Play an essential important role	vp	active VP	ideational	description
186	Play an important role	vp	active VP	ideational	description
187	Prefer to lead	vp	active VP	ideational	action
188	Put out	vp	active VP	ideational	action
189	Put us in danger	vp	active VP	ideational	action
190	Quarrel for	vp	active VP	ideational	action
191	Regard it as	vp	active VP	ideational	action
192	Remind us of	vp	active VP	ideational	action
193	Ride on swing	vp	active VP	ideational	action
194	Right the wrong	vp	active VP	ideational	action
195	Seek fortune	vp	active VP	ideational	action
196	Set fire for food	vp	active VP	ideational	action
197	Set fire to	vp	active VP	ideational	action
198	Set my mind to	vp	active VP	interpersonal	stance
199	Show great concern to	vp	active VP	ideational	action
200	Show his respect to	vp	active VP	ideational	action
201	Spare our time enjoying	vp	active VP	ideational	action
202	Spring plough	vp	active VP	ideational	action
203	Start from	vp	active VP	ideational	action
204	Stay in touch	vp	active VP	ideational	action
205	Stop blaming	vp	active VP	ideational	action
206	Stumble across	vp	active VP	ideational	action
207	Succeed in	vp	active VP	ideational	action
208	Take a breath	vp	active VP	ideational	action
209	Take on a new look	vp	active VP	ideational	action
210	Take spring outing	vp	active VP	ideational	action
211	Taking photographs	vp	active VP	ideational	action
212	Tell apart	vp	active VP	ideational	action

213	Told…not to do	vp	active VP	ideational	action
214	Transfer into	vp	active VP	ideational	action
215	Transform into	vp	active VP	ideational	action
216	Treasure it as	vp	active VP	ideational	action
217	Treat…as…	vp	active VP	ideational	action
218	Used to	vp	active VP	ideational	description
219	Wear willow branches on the head	vp	active VP	ideational	action
220	Wrapped for burial	vp	active VP	ideational	description

Appendix 13 Structural and Functional categorisation of four-word clusters

Year 1 Structural and Functional categorisation of four-word clusters

N	cluster	structure	sub-structure	function	subfunction
1	the dragon boat festival	np	other np	ideational	description
2	dragon boat festival is	vp	pronoun/np+be …	ideational	description
3	one of the most	np	np+of	ideational	quantification
4	is one of the	vp	copula be+np/adj.p	ideational	quantification
5	of the most important	pp	pp+of	textual	transition
6	festival is one of	vp	pronoun/np+be …	ideational	quantification
7	boat festival is one	vp	pronoun/np+be …	ideational	description
8	the most important traditional	np	other NP	ideational	description
9	I was born in	vp	pronoun/np+be …	ideational	description
10	important traditional festivals in	np	np+other post modifier	ideational	description
11	most important traditional festivals	np	other NP	ideational	description
12	more and more people	np	other NP	textual	transition
13	my English name is	vp	pronoun/np+be …	ideational	description
14	traditional festivals in china	np	np+other post modifier	ideational	description
15	boat festival is a	vp	pronoun/np+be …	ideational	description
16	I want to be	vp	active vp	interpersonal	stance
17	as far as I	pp	other pp	textual	framing
18	that the dragon boat	np	other NP	ideational	description
19	I went to the	vp	active vp	ideational	action
20	as is known to	clausal	as-clause	textual	generalisation
21	want to be a	vp	active vp	ideational	action
22	try my best to	vp	active vp	ideational	action
23	far as I am	other	other	textual	framing
24	there is no doubt	vp	pronoun/np+be …	interpersonal	stance
25	about the dragon boat	pp	other pp	ideational	description

26	Hubei university of automotive	np	np+of	ideational	description
27	is no doubt that	vp	copula be+np/adj.p	interpersonal	stance
28	of the dragon boat	pp	other pp	ideational	description
29	university of automotive technology	pp	pp+of	ideational	description
30	when it comes to	clausal	wh-clause	textual	framing
31	try our best to	vp	active vp	ideational	action
32	people who are in	np	np+other post modifier	ideational	description
33	and my English name	np	other NP	ideational	description
34	I think it is	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
35	dragon boat festival in	np	other NP	ideational	description
36	to help those in	clausal	to-clause	ideational	action
37	when I was six	clausal	wh-clause	ideational	description
38	so I want to	vp	active vp	textual	resultative
39	I was six years	vp	pronoun/np+be …	ideational	description
40	in the dragon boat	pp	other pp	ideational	location
41	jumped into the river	vp	active vp	ideational	action
42	it is said that	vp	pronoun/np+be …	textual	framing
43	who are in need	clausal	wh-clause	ideational	description
44	to be a volunteer	clausal	to-clause	ideational	action
45	help those in need	vp	active vp	ideational	action
46	if you want to	clausal	if-clause	interpersonal	engagement
47	to be a teacher	clausal	to-clause	ideational	action
48	his country very much	other	other	ideational	description
49	is international volunteers day	vp	copula be+np/adj.p	ideational	description
50	is not only a	vp	copula be+np/adj.p	textual	transition
51	it is a good	vp	pronoun/np+be …	interpersonal	stance
52	not only can we	other	template	textual	resultative
53	was born in a	vp	copula be+np/adj.p	ideational	description
54	when I was a	clausal	wh-clause	ideational	description
55	is the dragon boat	vp	copula be+np/adj.p	ideational	description

56	the benefits of volunteering	np	np+of	ideational	description
57	when I was in	clausal	wh-clause	ideational	description
58	I'll try my best	vp	active vp	interpersonal	stance
59	and more people are	vp	pronoun/np+be …	ideational	description
60	at the same time	pp	other pp	textual	transition
61	can learn a lot	vp	active vp	ideational	action
62	dragon boat festival to	np	np+other post modifier	ideational	description
63	he jumped into the	vp	active vp	ideational	description
64	I major in English	vp	active vp	ideational	description
65	I went to school	vp	active vp	ideational	description
66	into the Miluo river	pp	other pp	ideational	location
67	m an English major	vp	pronoun/np+be …	ideational	description
68	million registered volunteers who	np	np+other post modifier	ideational	quantification
69	my major is English	vp	pronoun/np+be …	ideational	description
70	my name is Wang	vp	pronoun/np+be …	ideational	description
71	people around the world	np	np+other post modifier	ideational	description
72	registered volunteers who have	np	np+other post modifier	ideational	description
73	the dragon boat competition	np	other NP	ideational	description
74	a volunteer is a	vp	pronoun/np+be …	ideational	description
75	celebrate the dragon boat	vp	active vp	ideational	action
76	in china it is	pp	pp+other post modifier	ideational	location
77	billion hours of volunteer	np	np+of	ideational	quantification
78	boat festival is an	vp	pronoun/np+be …	ideational	description
79	boat festival is the	vp	pronoun/np+be …	ideational	description
80	festival is a traditional	vp	pronoun/np+be …	ideational	description
81	festivals in china it	pp	other pp	ideational	location
82	have provided more than	vp	active vp	ideational	action
83	have volunteered to help	clausal	to-clause	ideational	action

84	hours of volunteer work	np	np+of	ideational	quantification
85	I come from class	vp	active vp	ideational	description
86	I hope I can	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
87	I went to a	vp	active vp	ideational	description
88	I went to primary	vp	active vp	ideational	description
89	important traditional festival in	np	np+other post modifier	ideational	description
90	into the river to	pp	other pp	ideational	location
91	is a good way	vp	copula be+np/adj.p	interpersonal	stance
92	is a traditional festival	vp	copula be+np/adj.p	ideational	description
93	is known to us	vp	copula be+np/adj.p	textual	generalisation
94	it is not only	vp	pronoun/np+be …	textual	transition
95	jumped into the Miluo	vp	active vp	ideational	action
96	loved his country very	vp	active vp	ideational	action
97	millions of people around	np	np+of	ideational	quantification
98	most important festival in	np	np+other post modifier	ideational	description
99	my dream is to	vp	pronoun/np+be …	ideational	description
100	observe the festival in	vp	active vp	ideational	action
101	of millions of people	np	np+of	ideational	quantification
102	of people around the	np	np+other post modifier	ideational	description
103	people observe the festival	vp	pronoun/np+be …	ideational	action
104	should try our best	vp	active vp	interpersonal	stance
105	tens of millions of	np	np+of	ideational	quantification
106	the first time i	np	np+other post modifier	ideational	location
107	the most important festival	np	other NP	ideational	description
108	the most important festivals	np	other NP	ideational	description
109	those who are in	np	np+other post modifier	ideational	description
110	to go to school	clausal	to-clause	ideational	action
111	to help those who	clausal	to-clause	ideational	action

112	to take part in	clausal	to-clause	ideational	action
113	to the primary school	pp	other pp	ideational	location
114	traditional festival in china	np	np+other post modifier	ideational	description
115	traditional festivals in our	np	np+other post modifier	ideational	description
116	volunteers who have provided	np	np+other post modifier	ideational	description
117	you can call me	vp	pronoun/np+be …	interpersonal	engagement
118	being a volunteer is	vp	copula be+np/adj.p	ideational	action
119	to observe this festival	clausal	to-clause	ideational	action
120	when the dragon boat	clausal	wh-clause	ideational	description
121	a lot of things	np	np+of	ideational	quantification
122	a traditional festival in	np	np+other post modifier	ideational	description
123	all over the world	pp	other pp	ideational	description
124	around the world have	pp	other pp	ideational	description
125	become more and more	vp	active vp	textual	transition
126	boat festival is not	vp	pronoun/np+be …	ideational	description
127	festival is a very	vp	pronoun/np+be …	ideational	description
128	festival is not only	vp	pronoun/np+be …	ideational	description
129	festivals in our country	np	np+other post modifier	ideational	description
130	help those who are	vp	active vp	ideational	action
131	I was five years	vp	pronoun/np+be …	ideational	description
132	I was sent to	vp	pronoun/np+be …	ideational	description
133	in memory of the	pp	pp+of	ideational	description
134	is in memory of	vp	copula be+np/adj.p	ideational	description
135	is my English name	vp	copula be+np/adj.p	ideational	description
136	learn a lot of	vp	active vp	ideational	action
137	like to make friends	clausal	to-clause	ideational	action
138	more and more popular	other	other	textual	transition
139	most important traditional festival	np	other NP	ideational	description

140	necessary for us to	adj.p/adv.p	adj.p	interpersonal	stance
141	on the dragon boat	pp	other pp	ideational	location
142	our best to help	np	np+other post modifier	ideational	description
143	school when I was	clausal	wh-clause	ideational	description
144	should pay more attention	vp	active vp	ideational	action
145	the hometown of Qu	np	np+of	ideational	description
146	the world have volunteered	vp	active vp	ideational	description
147	to make friends with	clausal	to-clause	ideational	action
148	to observe the festival	clausal	to-clause	ideational	action
149	to the middle school	pp	other pp	ideational	location
150	volunteered to help those	clausal	to-clause	ideational	action
151	we should pay more	vp	active vp	interpersonal	stance
152	we should try our	vp	active vp	interpersonal	stance
153	went to primary school	vp	active vp	ideational	action
154	when I was five	clausal	wh-clause	ideational	description
155	which is said to	clausal	wh-clause	textual	transition
156	who have provided more	clausal	wh-clause	ideational	description
157	with the development of	pp	pp+of	ideational	description
158	world have volunteered to	vp	active vp	ideational	action

Year 3 Structural and Functional categorisation of four-word clusters

N	Word	**Structure**	**sub-structure**	**function**	**sub-function**
1	the Qing Ming festival	np	other np	ideational	description
2	Qing Ming festival is	vp	pronoun/np+be	ideational	description
3	we should help strangers	vp	active vp	interpersonal	stance
4	more and more people	np	other np	textual	transition
5	that we should help	clausal	that-clause	interpersonal	stance
6	should we help strangers	vp	active vp	interpersonal	stance

7	is one of the	vp	copula be+np/adj.p	ideational	quantification
8	as far as i	pp	other pp	textual	framing
9	with the development of	pp	pp+of	ideational	description
10	willing to read classics	clausal	to-clause	ideational	description
11	far as i am	other	other	textual	framing
12	the importance of reading	np	np+of	ideational	description
13	people are willing to	vp	pronoun/np+be	ideational	description
14	is very important for	vp	copula be+np/adj.p	interpersonal	stance
15	are willing to read	vp	copula be+np/adj.p	ideational	description
16	festival is one of	vp	pronoun/np+be	ideational	quantification
17	try our best to	clausal	to-clause	ideational	action
18	importance of reading classics	np	np+of	ideational	description
19	reading classics is very	vp	pronoun/np+be	ideational	description
20	classics is very important	vp	pronoun/np+be	interpersonal	stance
21	Ming festival is a	vp	pronoun/np+be	ideational	description
22	tomb sweeping day is	vp	pronoun/np+be	ideational	description
23	less and less people	np	other np	textual	transition
24	Ming festival is one	vp	pronoun/np+be	ideational	description
25	our best to help	np	np+other post modifier	ideational	action
26	and more people are	vp	pronoun/np+be	ideational	description
27	is not only a	vp	copula be+np/adj.p	textual	transition
28	is the most important	vp	copula be+np/adj.p	interpersonal	stance
29	Qing Ming is a	vp	pronoun/np+be	ideational	description
30	I think we should	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
31	we should try our	vp	active vp	interpersonal	stance
32	is a time to	vp	copula be+np/adj.p	ideational	description
33	for us to read	pp	pp+other post modifer	ideational	description
34	a combination of sadness	np	np+of	ideational	quantification
35	as is known to	clausal	as-clause	textual	generalisation
36	combination of sadness and	np	np+of	ideational	quantification

37	seasonal division points in	np	np+other post modifier	ideational	location
38	should try our best	vp	active vp	interpersonal	stance
39	the temperature will rise	vp	active vp	ideational	action
40	give a hand to	vp	active vp	ideational	action
41	is no doubt that	vp	copula be+np/adj.p	interpersonal	stance
42	a period to honour	np	np+other post modifier	ideational	location
43	ancestors and sweep the	np	np+other post modifier	ideational	description
44	bow before the memorial	vp	active vp	ideational	action
45	each year after the	np	np+other post modifier	ideational	location
46	helping others is a	vp	pronoun/np+be	ideational	description
47	is the high time	vp	copula be+np/adj.p	ideational	description
48	it is the high	vp	copula be+np/adj.p	ideational	description
49	learn to protect ourselves	vp	active vp	ideational	action
50	less people are willing	vp	pronoun/np+be	ideational	description
51	of sadness and happiness	pp	other pp	ideational	description
52	one of the most	np	np+of	ideational	quantification
53	the high time for	np	np+other post modifier	ideational	location
54	to read classics as	clausal	to-clause	ideational	action
55	when it comes to	clausal	wh-clause	textual	framing
56	take advantage of our	vp	active vp	ideational	action
57	there is no doubt	vp	pronoun/np+be	interpersonal	stance
58	a festival of commemoration	np	np+of	ideational	description
59	a time to remember	np	np+other post modifier	ideational	location
60	and less people are	vp	pronoun/np+be	ideational	description
61	and sweep the tombs	vp	active vp	ideational	action
62	contrast to the sadness	np	other np	textual	transition
63	extended the day to	vp	active vp	ideational	action
64	festival sees a combination	vp	active vp	ideational	description
65	in contrast to the	pp	other pp	textual	transition
66	in my point of	pp	pp+of	textual	framing

67	is a period to	vp	copula be+np/adj.p	ideational	description
68	make our society more	vp	active vp	ideational	action
69	Ming festival is also	vp	pronoun/np+be	ideational	description
70	Ming festival sees a	vp	active vp	ideational	description
71	no doubt that we	clausal	that-clause	interpersonal	stance
72	period to honour and	np	np+other post modifier	ideational	location
73	Qing Ming festival as	np	other np	ideational	description
74	Qing Ming festival sees	vp	active vp	ideational	description
75	rise up and rainfall	vp	active vp	ideational	action
76	sadness of the tomb	np	np+of	ideational	description
77	sweep the tombs of	vp	active vp	ideational	action
78	temperature will rise up	vp	active vp	ideational	description
79	the development of the	np	np+of	ideational	description
80	the sadness of the	np	np+other post modifier	ideational	description
81	their ancestors and sweep	np	other np	ideational	description
82	this is the most	vp	pronoun/np+be	textual	transition
83	time to read classics	np	np+other post modifier	ideational	location
84	time to remember the	np	np+other post modifier	ideational	location
85	to read classics the	clausal	to-clause	ideational	action
86	to remember the dead	clausal	to-clause	ideational	action
87	to the sadness of	pp	pp+of	ideational	description
88	to their ancestors and	pp	other pp	ideational	description
89	will rise up and	vp	active vp	ideational	action
90	festival is a time	vp	pronoun/np+be	ideational	description
91	of food and spirit	pp	other pp	ideational	description
92	very important for us	adj.p/adv.p	adj.p	interpersonal	stance
93	a seasonal point to	np	np+other post modifier	ideational	location
94	a time to plant	np	np+other post modifier	ideational	location
95	advantage of our sympathy	np	np+of	ideational	description
96	ancestors often extended the	vp	active vp	ideational	description
97	and bow before the	vp	active vp	ideational	action

98	and only cold food	np	other np	ideational	description
99	and paper money and	np	other np	ideational	description
100	and rainfall increases it	np	other np	ideational	description
101	as our ancestors often	clausal	as-clause	ideational	description
102	at the same time	pp	other pp	textual	transition
103	be very cautious of	vp	copula be+np/adj.p	interpersonal	stance
104	before the memorial tablet	pp	other pp	ideational	location
105	burn incense and paper	vp	active vp	ideational	action
106	but the Qing Ming	np	other np	textual	transition
107	classics as a college	np	other np	ideational	description
108	cold food is served	vp	pronoun/np+be	ideational	description
109	commemoration the Qing Ming	np	other np	ideational	description
110	cook on this day	vp	active vp	ideational	action
111	day and only cold	np	other np	ideational	description
112	festival as our ancestors	np	other np	ideational	description
113	festival is not only	vp	pronoun/np+be	ideational	description
114	festival of commemoration the	np	np+of	ideational	description
115	have little time to	vp	active vp	ideational	quantification
116	helping others can sometimes	vp	active vp	ideational	action
117	helping others is helping	vp	pronoun/np+be	ideational	description
118	helping strangers is a	vp	pronoun/np+be	ideational	description
119	high time for spring	np	np+other post modifier	ideational	location
120	hope of spring on	np	np+of	ideational	description
121	important day of sacrifice	np	np+of	ideational	description
122	incense and paper money	np	other np	ideational	description
123	increases it is the	vp	active vp	ideational	action
124	is more a festival	vp	copula be+np/adj.p	ideational	description
125	it is said that	vp	pronoun/np+be	textual	framing
126	Ming festival is not	vp	pronoun/np+be	ideational	description
127	money and bow before	np	other np	ideational	description
128	more a festival of	np	np+of	ideational	description

129	most important day of	np	np+of	ideational	description
130	not cook on this	vp	active vp	ideational	action
131	not only a seasonal	other	template	textual	transition
132	of commemoration the Qing	pp	other pp	ideational	description
133	of spring on this	pp	other pp	ideational	location
134	of the most important	pp	other pp	textual	transition
135	offer sacrifices to their	vp	active vp	ideational	action
136	often extended the day	vp	active vp	ideational	action
137	on the way to	pp	pp+other post modifer	ideational	description
138	on this day and	pp	other pp	ideational	location
139	only a seasonal point	np	other np	ideational	description
140	only cold food is	vp	pronoun/np+be	ideational	description
141	our ancestors often extended	vp	active vp	ideational	description
142	our society more harmonious	np	other np	ideational	description
143	paper money and bow	np	other np	ideational	description
144	point to guide farm	np	np+other post modifier	ideational	description
145	Qing Ming festival was	vp	pronoun/np+be	ideational	description
146	rainfall increases it is	vp	active vp	ideational	description
147	reading classics we can	vp	active vp	interpersonal	stance
148	remember the dead and	vp	active vp	ideational	action
149	sacrifices to their ancestors	vp	active vp	ideational	action
150	seasonal point to guide	np	np+other post modifier	ideational	location
151	sees a combination of	vp	active vp	ideational	action
152	so that we can	other	template	textual	resultative
153	sowing but the Qing	np	other np	ideational	description
154	spirits of deceased ancestors	np	np+of	ideational	description
155	spring on this day	pp	other pp	ideational	location
156	tend to be very	clausal	to-clause	textual	resultative
157	that helping others can	clausal	that-clause	interpersonal	stance

158	the beginning of the	np	np+of	ideational	location
159	the dead and the	np	other np	ideational	description
160	the most important day	np	other np	ideational	description
161	the spirits of deceased	np	np+of	ideational	description
162	the tombs of the	np	np+of	ideational	description
163	they were later combined	vp	pronoun/np+be	ideational	description
164	they will not cook	vp	pronoun/np+be	ideational	action
165	this day and only	np	other np	ideational	description
166	to be very cautious	clausal	to-clause	ideational	description
167	to give a hand	clausal	to-clause	ideational	action
168	to help strangers in	clausal	to-clause	ideational	description
169	we should help others	vp	active vp	interpersonal	stance
170	we should try to	vp	active vp	interpersonal	stance
171	when the lord succeeded	clausal	wh-clause	ideational	description
172	when you are in	clausal	wh-clause	ideational	description
173	whether we should help	other	template	textual	transition
174	will not cook on	vp	active vp	ideational	action
175	fewer and fewer people	np	other np	textual	transition
176	food and spirit money	np	other np	ideational	description
177	a piece of his	np	np+of	ideational	quantification
178	also a time to	np	np+other post modifier	ideational	description
179	also learn to protect	vp	active vp	ideational	action
180	and favourites of the	np	np+of	ideational	description
181	and happiness this is	other	other	ideational	description
182	and minority ethnic groups	np	other np	ideational	description
183	and pay respect to	vp	active vp	ideational	action
184	and sowing but the	np	other np	ideational	description
185	and trees grow fast	vp	active vp	ideational	description
186	at this time offer	pp	other pp	ideational	location
187	become more and more	vp	active vp	textual	transition
188	becoming more and more	vp	active vp	textual	transition
189	becoming the ruler of	vp	active vp	ideational	action
190	before the Qing Ming	pp	other pp	ideational	location

191	both the Han and	np	other np	ideational	description
192	by serving a piece	pp	other pp	ideational	description
193	day after the winter	np	other np	ideational	location
194	day before the Qing	np	other np	ideational	location
195	doubt that we should	clausal	that-clause	interpersonal	stance
196	festival was called arbour	vp	pronoun/np+be	ideational	description
197	festival was usually one	vp	pronoun/np+be	ideational	description
198	flowers and favourites of	np	other np	ideational	description
199	from my point of	pp	pp+of	textual	framing
200	go to the library	vp	active vp	ideational	action
201	growing number of people	np	np+of	ideational	quantification
202	Han and minority ethnic	np	other np	ideational	description
203	happiness this is the	np	other np	ideational	description
204	has always been considered	vp	passive verb	textual	generalisation
205	have a lot of	vp	active vp	ideational	quantification
206	high and trees grow	other	other	ideational	description
207	important for us to	adj.p/adv.p	adj.p	interpersonal	stance
208	in becoming the ruler	pp	other pp	ideational	description
209	in our daily life	pp	other pp	ideational	description
210	in Shanxi province in	pp	other pp	ideational	location
211	is high and trees	vp	copula be+np/adj.p	ideational	description
212	is very important to	vp	copula be+np/adj.p	interpersonal	stance
213	it is a period	vp	pronoun/np+be	ideational	description
214	it is a time	vp	pronoun/np+be	ideational	description
215	it is more a	vp	pronoun/np+be	ideational	description
216	know the importance of	vp	active vp	ideational	action
217	leg when the lord	clausal	wh-clause	ideational	description
218	lived in Shanxi province	vp	active vp	ideational	location
219	lord ordered all fires	vp	active vp	ideational	action
220	lord succeeded in becoming	vp	active vp	ideational	action
221	love to fly kites	vp	active vp	ideational	action
222	maybe the best of	np	np+of	interpersonal	stance
223	Ming festival as our	np	np+other post modifier	ideational	description

224	Ming festival was called	np	np+other post modifier	ideational	description
225	Ming is a time	vp	pronoun/np+be	ideational	description
226	minority ethnic groups at	np	other np	ideational	location
227	money could keep them	vp	active vp	ideational	description
228	no choice but to	clausal	to-clause	textual	resultative
229	on the anniversary of	pp	pp+of	ideational	location
230	one day before the	np	other np	ideational	location
231	ordered all fires in	vp	active vp	ideational	action
232	out on the anniversary	np	other np	ideational	location
233	pay more attention to	vp	active vp	ideational	action
234	people are afraid of	vp	pronoun/np+be	ideational	description
235	piece of his own	np	np+of	ideational	description
236	protect ourselves from being	vp	active vp	ideational	action
237	remain where he was	clausal	wh-clause	ideational	location
238	s life by serving	np	other np	ideational	action
239	s no doubt that	vp	copula be+np/adj.p	interpersonal	stance
240	sadness and happiness this	np	other np	ideational	description
241	serving a piece of	vp	active vp	ideational	action
242	should also learn to	clausal	to-clause	ideational	action
243	succeeded in becoming the	vp	active vp	ideational	action
244	sweep tombs and offer	vp	active vp	ideational	action
245	that the spirits of	np	np+of	ideational	description
246	the custom of spring	np	np+of	ideational	description
247	the Han and minority	np	np+of	ideational	description
248	the high pace of	np	np+of	ideational	quantification
249	the lord ordered all	vp	active vp	ideational	action
250	the lord succeeded in	vp	active vp	ideational	action
251	the memorial tablet in	np	np+other post modifier	ideational	description
252	the survival rate of	np	np+of	ideational	quantification
253	the tomb sweeping day	np	other np	ideational	description
254	the trees and grass	np	other np	ideational	description

255	think reading classics is	vp	active vp	ideational	action
256	time offer sacrifices to	np	other np	ideational	action
257	to death to commemorate	clausal	to-clause	ideational	action
258	to fly kites during	clausal	to-clause	ideational	action
259	to pay respect to	clausal	to-clause	ideational	action
260	to protect ourselves from	clausal	to-clause	ideational	action
261	to read classics and	clausal	to-clause	ideational	action
262	to read the classics	clausal	to-clause	ideational	action
263	to remain where he	clausal	to-clause	ideational	location
264	to sweep tombs and	clausal	to-clause	ideational	action
265	tombs and offer sacrifices	np	other np	ideational	description
266	tombs of the diseased	np	other np	ideational	description
267	up and rainfall increases	vp	active vp	ideational	description
268	usually one day before	np	other np	ideational	location
269	was usually one day	vp	copula be+np/adj.p	ideational	description
270	we should also learn	vp	active vp	interpersonal	stance
271	where he was and	clausal	wh-clause	ideational	location
272	who are in trouble	clausal	wh-clause	ideational	description
273	who are willing to	clausal	wh-clause	ideational	description
274	who lived in Shanxi	clausal	wh-clause	ideational	description
275	who take advantage of	clausal	wh-clause	ideational	description
276	you are in trouble	vp	pronoun/np+be	interpersonal	engagement

Appendix 14 Shared formulaic language

Shared formulaic strings in the Year 1 and Year 3 corpora

No.	strings	structure	sub-structure	function	sub-function
1	According to	pp	pp+ other post modifiers	textual	resultative
	***According to legends[13]	pp	pp+ other post modifiers	textual	resultative
2	As a result	PP	other pp	textual	resultative
	***As a result	pp	other pp	textual	resultative
3	As far as I am concerned	pp	other pp	textual	framing
	***As far as I am concerned	pp	other pp	textual	framing
4	As it is known to us all	clausal	as-clause	textual	generalisation
	***As it is known	clausal	as-clause	textual	generalisation
5	At the same time	pp	other pp	textual	transition
	***At the same time	pp	other pp	textual	transition
6	First…Second…Finally	other	template	textual	structuring
	***First…Second…Finally	other	template	textual	structuring
7	For example	pp	other pp	textual	exemplification
	***For example	pp	other pp	textual	exemplification
8	From my point of view	pp	pp+of	textual	framing
	***From my point of view	pp	pp+of	textual	framing
9	I believe that	clausal	1st/2nd person pronoun+dep. cl.	interpersonal	stance
	***I firmly believe that we	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
10	In order to	pp	pp+ other post modifiers	textual	resultative
	***In order to	pp	pp+ other post modifiers	textual	resultative
11	In my point of view	pp	pp+of	textual	framing
	***From my point of view	pp	pp+of	textual	framing
12	No pain, no gain	other	formulae	interpersonal	stance
	***no pain, no gain	other	formulae	interpersonal	stance
13	Not only…but also	other	template	textual	transition
	***Not only…but also…	other	template	textual	transition
14	On the other hand	pp	other pp	textual	transition

[13] Items in the list marked with *** are from the Year 3 corpora.

	***On the other hand	pp	other pp	textual	transition
15	One of the	np	np+of	ideational	quantification
	***One of the most important	np	np+of	ideational	quantification
16	Play an essential important role	vp	active vp	ideational	description
	***Play an important role	vp	active vp	ideational	description
17	Prefer to	vp	active vp	ideational	action
	***Prefer to lead	vp	active vp	ideational	action
18	To begin with	vp	active vp	textual	structuring
	***To begin with…then…at last	other	template	textual	structuring
19	Try our best	vp	active vp	ideational	action
	***Try one's best to	clausal	to-clause	ideational	action
20	Used to be	vp	active vp	ideational	description
	***Used to	vp	active vp	ideational	description
21	What's more	other	ill-formed	textual	transition
	***What's more	other	ill-formed	textual	transition

Shared four-word clusters in the Year 1 and Year 3 corpora

No.	cluster	structure	sub-structure	function	Sub-function
1	and more people are	vp	pronoun/np+be …	ideational	description
	***and more people are	vp	pronoun/np+be	ideational	description
2	as far as I	pp	other pp	textual	framing
	***as far as i	pp	other pp	textual	framing
3	as is known to	clausal	as-clause	textual	generalisation
	***as is known to	clausal	as-clause	textual	generalisation
4	at the same time	pp	other pp	textual	transition
	***at the same time	pp	other pp	textual	transition
5	become more and more	vp	active vp	textual	transition
	***become more and more	vp	active vp	textual	transition
6	far as I am	other	other	textual	framing
	***far as i am	other	other	textual	framing
7	festival is not only	vp	pronoun/np+be …	ideational	description
	***festival is not only	vp	pronoun/np+be	ideational	description
8	festival is one of	vp	pronoun/np+be …	ideational	quantification
	***festival is one of	vp	pronoun/np+be	ideational	quantification
9	is not only a	vp	copula be+np/adj.p	textual	transition

	***is not only a	vp	copula be+np/adj.p	textual	transition
10	is one of the	vp	copula be+np/adj.p	ideational	quantification
	***is one of the	vp	copula be+np/adj.p	ideational	quantification
11	it is said that	vp	pronoun/np+be …	textual	framing
	***it is said that	vp	pronoun/np+be …	textual	framing
12	more and more people	np	other NP	textual	transition
	***more and more people	np	other NP	textual	transition
13	of the most important	pp	pp+of	textual	transition
	***of the most important	pp	pp+of	textual	transition
14	one of the most	np	np+of	ideational	quantification
	***one of the most	np	np+of	ideational	quantification
15	our best to help	np	np+other post modifier	ideational	description
	***our best to help	np	np+other post modifier	ideational	description
16	should try our best	vp	active vp	interpersonal	stance
	***should try our best	vp	active vp	interpersonal	stance
17	there is no doubt	vp	pronoun/np+be …	interpersonal	stance
	***there is no doubt	vp	pronoun/np+be …	interpersonal	stance
18	we should try our	vp	active vp	interpersonal	stance
	***we should try our	vp	active vp	interpersonal	stance
19	when it comes to	clausal	wh-clause	textual	framing
	***when it comes to	clausal	wh-clause	textual	framing
20	with the development of	pp	pp+of	ideational	description
	***with the development of	pp	pp+of	ideational	description

Core formulaic language in the Year 1 corpus

No.	**strings**	**structure**	**sub-structure**	**function**	**sub-function**
1	as far as I	PP	other pp	textual	framing
	**As far as I am concerned[14]	PP	other pp	textual	framing

[14] Items in the list marked with ** are student identified formulaic strings

2	as is known to	clausal	as-clause	textual	generalisation
	**As is known to all	clausal	as-clause	textual	generalisation
3	At the same time	pp	other pp	textual	transition
	**at the same time	pp	other pp	textual	transition
4	I hope I can	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
	**I hope I'll	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
5	I think it is	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
	**I think it's a	clausal	1st/2nd person pronoun+dep.cl.	interpersonal	stance
6	I'll try my best	vp	active vp	interpersonal	stance
	**I'll try my best to be impressive	vp	pronoun/np+be…	interpersonal	stance
7	if you want to	clausal	if-clause	interpersonal	engagement
	**If you want to be loved, love others.	clausal	if-clause	interpersonal	engagement
8	On the other hand	pp	other pp	textual	transition
	**On the one hand…on the other hand	other	template	textual	transition
9	One of the most	np	np+of	ideational	quantification
	**one of the	np	np+of	ideational	quantification
10	The Dragon boat festival	np	other np	ideational	description
	**the dragon boat festival	np	other np	ideational	description
11	those who are in	np	np+other post modifier	ideational	description
	**those who are needed help	np	np+ other postmodifier	ideational	description
12	Try our best to	vp	active vp	ideational	action
	**try our best	vp	active vp	ideational	action
13	When it comes to	clausal	wh-clause	textual	structuring
	**when it comes to	clausal	wh-clause	textual	structuring
14	which is said to	clausal	wh-clause	textual	transition
	**Which was said to	clausal	wh-clause	textual	transition
15	you can call me	vp	active vp	interpersonal	engagement
	**you can call me	vp	active vp	interpersonal	engagement

Core formulaic language in the Year 3 corpora

No.	**strings**	**structure**	**sub-structure**	**function**	**Sub-function**
1	as far as I	pp	other pp	textual	framing
	**As far as I am concerned	pp	other pp	textual	framing

2	At the same time	pp	other pp	textual	transition
	**at the same time	pp	other pp	textual	transition
3	from my point of	pp	pp+of	textual	framing
	**From my point of view	pp	pp+of	textual	framing
4	helping others is helping	vp	pronoun/np+be	ideational	description
	**Help others is help ourselves	other	formulae	interpersonal	stance
5	in my point of	pp	pp+of	textual	framing
	**In my point of view	pp	pp+of	textual	framing
6	it is said that	vp	pronoun/np+be	textual	framing
	**It's said	vp	pronoun/np+be…	textual	framing
7	more and more people	np	other np	textual	transition
	**More and more	Adj.P /Adv.p	adv.p	textual	transition
8	not only a seasonal	other	template	textual	transition
	**Not only…but also…	clausal	other	textual	transition
9	one of the most	np	np+of	ideational	quantification
	**One of the most important	np	np+of	ideational	quantification
10	pay more attention to	vp	active vp	ideational	action
	**Pay attention to	vp	active vp	ideational	action
11	people are willing to	vp	pronoun/np+be	ideational	description
	**People are willing to go out	vp	pronoun/np+be	ideational	description
12	the Qing Ming festival	np	other np	ideational	description
	**Qing Ming Festival	np	other np	ideational	description
13	the tomb sweeping day	np	other np	ideational	description
	**Tomb sweeping day	np	other np	ideational	description
14	there is no doubt	vp	pronoun/np+be	interpersonal	stance
	**There is no doubt that	vp	pronoun/np+be…	interpersonal	stance
15	try our best to	clausal	to-clause	ideational	action
	**Try one's best to	clausal	to-clause	ideational	action

Appendix 15 Selected examples and translation for the interview analysis

Example 1

1. “Fixed collocations, fixed phrases, and fixed sentence structures (INV 003)”.
   固定的搭配啊，固定的词组啊，固定的句型啊 (INV 003)。
2. “Since he emphasised a lot about the fixed phrases, it might be formulaic language… fixed phrases, fixed sentence structure, and some famous quotes or proverbs (INV 004)”.
   因为就是他比较强调固定短语，可能就是语块吧……固定短语，固定句型，还有名人名言 (INV 004) 。

Example 2

1. “In general, the first paragraph will present the main point, take the writing “the part-time job” as an example. I will present my viewpoint towards the part time job first, and then in the second paragraph, I will talk about opinions from other people. After this, I will focus on my own point. Finally, in the fourth paragraph, it is the conclusion (INV 004)”.
   基本上是，要点。基本上第一段，以做兼职为例吧，比如说先摆出做兼职这个观点，第二段写有人对它的看法，另一些人对它的看法，第三段写我个人的看法，然后最后再结尾 (INV 004)。
2. “Usually in the exam, the time is limited, so I do not write down any outline. I know I have to write the three-paragraph composition for the argumentative essay (INV 007)”.
   因为一般考试嘛，时间比较紧张，所以不会列提纲，议论文就三段式嘛 (INV 007)。

Example 3

1. “For example, a collocation or sentence structure, appear again and again (INV 002)”.
   比如说某一个单词词组或者是句型，重复出现 (INV 002)
2. “Some collocations with high frequency in use (INV 003)”.
   有一些高频率出现的词组的话 (INV 003)。

Example 4

1. “Some Chinese proverbs, for example, no pain, no gain. Other proverbs, like I can use in this writing, the fragrance always stays in the hand that gives the rose (INV 007)”.
   比如说中国的一些谚语吧，反正像 no pain, no gain. 比如，这次的作文题目，送人玫瑰，手有余香(INV 007)。
2. “When it comes to the proverbs, I think the most frequent one we have come across is where there is a will, there is a way (INV 010)”.
   如果谚语的话，我们接触的最多的就是 where there is a will, there is a way (INV 010)。

Example 5

1. "Fixed expressions, for example, be made up of, be composed of, and some others like the conjunctional words or phrases, like on the one hand…, on the other hand…; also, slang is formulaic language. I forget any example for that. I know less slang (INV 009)”.
   比如固定的表达吧，be made up of, be composed of, 还有那么关联词，像 on the one hand…, On the other hand…, 还有俚语那之类的，忘了，相对少一点 (INV 009) 。
2. “The greetings are formulaic language, which you used to communication in speaking every day (INV 010)”.

问候语的话就平时与别人用口语交流的话(INV 010)。

Example 6

"Some collocations consist of several words, and you never know its meaning until you put that in the text, within the context of the language use, you can read and understand the content back and forth, and you can get the meaning as it is in the text (INV 011)".

很多单个的搭配如果单独的去看根本不知道是什么意思，放在文中，你都不知道它讲的是什么，但是如果你知道了联系上下文，看了一个整体的段落，就知道它是什么意思了 (INV 011)。

Example 7

1. "The writing templates could help structure my writing correctly and in accordance with the requirements in the writing tasks and guarantee the better score (IVN 003)".
   根据那些标准的话，就是说用了模版，拿的分数才会好一点(IVN 003)。
2. "The writing templates help confine my language used in English writing within or following the requirement and pattern that the markers preferred (INV 005)".
   其实你能用的话，最好用。毕竟他们在改卷的时候，你要适应他的思维模式(INV 005)。

Example 8

1. If you can remember a template for English writing, and you can use some sentences in this templates directly, it is very handy, convenient, and save your time (INV 004)".
   让你记住一个模板，然后在提到一句话里与这个模板有关句型的时候，可以信手拈来，不费力这样比较节省时间 (INV 004)。
2. "When you write in English, you can remember these fixed items. When you need them, they just jump out in your mind, and you can use that in your writing. In this way, you can save time (INV 005)".
   你写英语的话，你记一些固定的这种东西，用的上的时候就直接蹦出来了，你就可以写，可以用，节省时间(INV 005)。

Example 9

1. "It makes your language more vivid, and more native-like (INV 001)".
   就是使你的语言更加有张力。然后就是更地道 (INV 001)。
2. "I think formulaic language is a great addition to my English. Suppose there are two compositions: One is without using any formulaic language, just a random writing, quite plain. The other one is with beautiful phrases, and formulaic language is used in a creative way in the text. It looks coherent and cohesive. Thus, the second compositions will have the higher mark and it makes the reader feel better (INV 003)".
   我觉得是锦上添花。譬如说有两份作文，一份里面的话，就没怎么用一些语块，就是随随便便，平平淡淡的一篇作文。然后另外一篇呢， 那个词组用的很漂亮，各个地方用的恰到好处，用法新颖，看起来，看起来很连贯，一气呵成的话，这个作文的话得分就会明显的高一点。给人的感觉就挺好的 (INV 003)。

3. “In the TEM-4 exam, for example, if you use formulaic language frequently in different ways, it will be a great addition to your language and you mark will be increased. That’s direct (INV 009)”.
   比如说你要考专四，你要是使用多次这些不同的语块，它就会给你的作文添彩，加分。这是最直接的(INV 009)。

Example 10

1. “In English, they have very funny phenomena, when you use idiom (INV 001)”.
2. “Since lots of funny things could be expressed in a simple way (INV 001)”.
3. “Formulaic language is funny, fragmental and figurative”.
   是感觉很好玩。很小，很形象(INV 011)。

Example 11

1. “It confines my language with lots of standards (INV 002)”,
   条条框框的束缚(INV 002)。
2. “I do not think it is good. There is too much limitation, just like writing a “Ba Gu Wen”, for example, this writing task. Also, I read a lot sample writing from the TEM-4, they all focused on the structures. Usually, when the teachers mark our English writing, they also check if you use language such as “first of all…second…third…”in your writing. I do not like this way of writing, as it is too rigid. I wish I can change it (INV 008)”.
   我觉得这样不好。这样就限制太大了，就像写八股文一样。比如我这篇文章就是，再比如我之前看的有的专四的作文就是，要求结构。但是一般老师改作文就要求结构，就看看你有没有 first of all…second…third…我就很不喜欢这种形式，觉得太死板了。我希望我能改变一下(INV 008)。

Example 12

“I feel that when we learned a new thing at first place, it was useful, anyway, we were not familiar with it. Yet, later on, I find that after you made some progress, you have to try to throw away this crutch finally. As they are the fixed templates, it is not good used by too many people (INV 010)”.
我觉得当我们刚接触一个新事物的时候，作用还是蛮大的。因为毕竟不是太了解，但是我觉得以后用的话，就是你英文水平进步之后，要尽量甩到掉这个拐杖嘛。因为毕竟是模板，用的人多了，总是不好(INV 010)。

Example 13

1. “The writing templates, I think it refers to a basic frame, or the way of thinking, it also includes some sentences with a certain structure (INV 002)”.
   固定文章模版，我感觉就是有那么一个基本框架，思路，句子的结构这些(INV 002)。
2. “In my opinion, in English writing, I usually use the writing templates, they generally are fixed. You then use those templates in your writing. You can change some parts in the templates. Thus I think it is the most importance one for me (INV 006)”.
   然后我觉得吧，写作嘛，基本上都是用那些个板块，基本上都是固定的，就是把模块用到写作中。就是把相应的词改变一下。所以我觉得它是最重要的(INV 006)。

Example 14

1. “The writing templates, I think it refers to a basic frame, or the way of thinking, it also includes some sentences with a certain structure. All of these enable you to get higher marks in the writing (INV 002)”.
   固定文章模版，我感觉就是有那么一个基本框架，思路，句子的结构这些都是得分点(INV 002)。
2. “If you want to survive the exams, you have to remember phrases, idioms and writing templates very well. If you just follow your own ideas, you do whatever you want, you would not get a good score (INV 006)”.
   觉得要应付考试的话肯定就是短语啊，俗语啊，固定文章模板这些肯定要记好。如果说按照自己的想法，想怎么写就怎么写的话，那样的分数肯定不会高 (INV 006)。

Example 15

1. “They can help you categorise your whole writing and present your ideas (INV 009)”.
   有了固定的文章模块你就能把他们结合起来就能把整篇文章就是表现出来，就是比较优美，比较耐读(INV 009)。
2. “They can make your writing clear to the reader (INV 007)”.
   用固定文章模版让人看起来会感觉很清晰(INV 007)。

Example 16

1. “Idioms and fixed expressions are language and cultural-specific, and contain rich connotation within them (INV 003)”.
   俗语和搭配语我觉得就是，我感觉这种的话他们就是有一些语言呀或者文化性的东西在里面。就是，很有内涵，就是不会太干巴巴的(INV 003).
2. “There are couples of idioms and fixed expressions or phrases in one sentence. They can give you a sense of the structure in a sentence (INV 009)”.
   像短语和俗语，短语，像一个句子的话，就是有很多短语及固定搭配构成的。就是有那种结构感(INV 009)。

Example 17

1. “People in the context of ESL used more idioms and fixed expressions than people in the context of EFL, like her, and she used them mostly in her writing rather than speaking (INV 003)”.
   我觉得在作文中运用俗语和固定搭配的话还比较多。然后作为外国人，或者讲英语国家的学生学英语，用的就比较多(INV 003)。
2. “I would use them in writing, since the larger repertoire you have in idioms and fixed expressions, the more vividness you can get in your language and the more fluency you can achieve when you are writing (INV 010)”.
   俗语及固定搭配的话，你写文章的话，自己懂得多的话，就可能比较增色，然后写下来的话也会比较流畅(INV 010)。

Example 18

1. “The phrases are the most basic ones. Since you learn the language to communicate with others, while, as a phrase, phrases that if you want to write something, if you want to speak, you cannot leave the phrases (INV 001)”.

2. “They are basic parts for constructing a sentence…for example, the combination of verb and noun. I don’t think it is a phrase. For instance, the expression, play table tennis, play is a verb, table tennis is a noun phrase. For me, I don’t think the expression play table tennis can be called a phrase. However, the teacher asked us to mark it and remember it as a phrase. For me, I think if you come across some expressions in your reading, you cannot say it is a phrase. (INV 004)”.
   短语就是基本上构成句子…。比如，动词短语和名词搭配，我个人认为他就不是短语。比如说，打乒乓球，play table tennis, 然后 play 是动词，table tennis 是名词，这个在我看来，他就不算短语，而老师提出来，叫我们记一下，作为短语，我个人认为就是，平常看书就能看到的，根本不算是什么短语 (INV 004)。

Example 19

1. “One interviewee reported that phrases are useful when you express meanings (INV 005)”.
   短语的话我觉得它的实用性比较强，比如你要表达什么意思的时候就会用到的(INV 005)。
2. Without the rules, you cannot get scores (INV 006)”.
   这些肯定要记好。如果说按照自己的想法，想怎么写就怎么写的话，那样的分数肯定不会高(INV 006)。

Example 20

1. “Using some phrases are not so familiar to their peers, but easy to be understood by audience/readers, which probably could draw attention from the marker or teacher, since the student wants to be different from others in the class, and show his/her language use is ingenious (INV 007)”.
   比如说有些很新鲜的短语，首先是别人看到这个短语很容易能够读懂，第二个就是，看的比较新颖。因为平常有些短语大家不用，如果你用在文章中，看起来比较新颖(INV 007)。
2. “They [classic phrases] just simplify the complex expressions, and people know it as it is popular, which makes your meaning expressed in a straightforward and impressive way (INV 012)”.
   因为有的像一句话，你可以用短语来表达。或者说在写文章中，如果有很多特别经典的短语的话，就会特别的出色(INV 012)。

Example 21

1. “Usually it is related to the impression that phatic expressions are not formal (INV 002)”.
   寒暄语在写作中出现的话就会显得不那么正式(INV 002)。
2. “Not appropriate used in writing, since writing is a formal way to express meaning (INV 004)”.
   问候语、口头禅和寒暄语。因为我觉得这些都不是特别地道，运用在作文中。因为作文嘛它比较正式，用在作文中显得不太得体(INV 004)。

Example 22

1. “Phatic expressions are not practical at all since she did not have many opportunities to communicate with foreigners, thus she barely used them in a real conversation (INV 011)”.

接下来不是很实用，因为在我们这种环境下没有跟外国人交谈的机会，所以口头禅问候语寒暄语用的不是很多(INV 011)。

2. “I feel I haven’t adapted myself in certain kinds of situation. For example, if some foreigners ask me “how are you”, I cannot really respond him immediately since I was thinking of the answer to “how am I”, so there is definite cultural difference. It is very strange if you think phatic expressions like how are you in Chinese. In Chinese, we prefer asked others like, “how is everything going recently?” “Have you had your meal?” “Where are you going?” Such kinds of questions. If I was asked suddenly, I feel weird. I think because I don’t get used to thinking in English and that’s it (INV 011)”.
   我觉得我现在还没有适应就是突然有个外国人跟你见面就问你 how are you，然后我就反应不过来，就是有很大的文化差异。但是如果你要是直接去想 how are you， 你就会感觉很怪异。基本上就是中国人问你，最近怎样，你去哪儿啊，你吃饭了没有，就是这样的。然后要是突然问就会觉得好怪异啊。就还是那种思维没有转变过来(INV 011)。

Example 23

1. “Compound is like the derivation, just a way of word formation (INV 004)”.
   因为合成词是单词构成的一种方法，比如说合成啊派生啊什么的(INV 004)。
2. “You can see these compound words. First, you can see the meaning of this word, because if you know the two of the words, which is consist of it (INV 001)”.

Example 24

1. “The compound words, for example, some very new ones, if you can use compound words in a novel way, which make your language polished and special (INV 007)”.
   合成词，因为平常有些大家不用，如果你用在文章中，看起来比较新颖(INV 007)。
2. “Compound words can enrich and enlarge your vocabulary quickly and efficiently (INV 012)”.
   合成词，它能帮助我们快速地记住很多单词嘛，单词面这个方面作用挺大的(INV 012)。

Example 25

“Terminologies, as each field has its own terminologies, such as science, social science, they have their subject-related terminologies. If you learn these terminologies, you will know the vocabulary of the whole area (INV 004)”.
专业术语，因为每一个领域的英语都有术语，比如说，科技类，社会类都有它的相应的术语，如果把每个领域的相应的基本术语学会，基本上这一学科的领域的那个单词就基本上能够过关了(INV 004) 。

Example 26

1. “Terminologies involve certain knowledge of the specific area, for example, international trading, that is quite a fixed way of saying things (INV 004)”.
   专业术语的话就涉及到固定的专业，比如说外贸，涉及到固定的表达。这些就是固定的领域(INV 009)。
2. “Terminology reduces the difficulties in the communication for both speaker and hearer, and makes all the interlocutors easier to be understood, especially within specific areas (INV 007)”.

所以说专业术语的话，不管你在哪个领域，对，你说的是专业术语，这样的话既方便别人能听懂，也更有利于你跟别人交流 (INV 007)。

Example 27

1. “Firstly, your saying is interesting; you have an interest in it, so you started to learn about it, and then when you master it you can use it in your life (INV 001)”.
2. “You have to memorise that, for example, this is a good article, then you memorise it, for you may use the language of it later on (INV 008)”.
   就是要背。这篇文章写得很好，要背一下，将来可能用得着(INV 008)。

Example 28

1. “So, you are kind of independent learning, and then you can learn something new, then put that into the practice, and then study the specific item deeper (INV 001)”.
2. “I think it is self-learning after the class. You learn it by yourself, and you read, then you can remember that. When it is mentioned by the teachers in the class, you will have a deeper impression of that. It’s like you acquire the knowledge of formulaic language and then practise that in the classroom (INV 004)”.
   我觉得还是私底下吧。因为私底下，看了，再把它记住，然后课堂上老师提出来的话， 印象会更深。就像那个在私底下获得知识，然后再课堂上得到实践 (INV 004)。

Example 29

1. “You have to understand the meaning of the language throughout, otherwise it is easy to make a mistake, and for example, it could be the surface meaning rather than the real meaning (INV 010)”.
   我觉得这个你要是对它有个很透彻的认识的话，就怕一知半解嘛。就容易错。可能就只是一个它很浅面的意思(INV 010)。

Example 30

1. “The model, I think it is quite imperative. The teacher asks you, you can only write the passage in three parts or five parts. The first part is about your common sense… and other parts are our description and the last part is a summary. The teacher told you that you have to do so. There is nothing have to speak (INV 001)”.
2. “The fixed writing templates is very important. Since only after we know very well about others writing, then we can imitate and use that (INV 012)”.
   固定的文章模板也挺重要，只有我们先了解了别人写的文章，然后才能去模仿和应用(INV 012)。

Example 31

1. “I always paid attention to the first three categories of formulaic language (writing templates, idioms and fixed expressions, and phrases); personally, I likes speech scripts, therefore, he especially focused and learned those kinds of formulaic language from speech scripts, and then read those items out repeatedly; I usually memorised them for later use in writing (INV 002)”.
   我比较偏向前三个。比如说我喜欢看演讲稿嘛，像奥巴马的，我就经常留意一些语言现象。而且还会反复的读，而且记下来，在写作中运用(INV 002)。

2. “Idioms and fixed expressions are important in our daily communication; the more items you have a good command of, the more helpful in the communication. Also, it took some time to remember, once you are familiar with those expressions, you can use them in convenience, it would better if you can put them in sentences or articles (INV 012)”.
   俗语和固定搭配也挺重要的，因为它是构成我们平时进行交流的很重要的部分。如果说我们掌握了很多的话，对我们在应用方面就是很重要。这些的话可能就是需要去花时间去背去记，这样的话用的就更熟练，顺手拈来就可以了。如果结合在一定的文章句子中就更好了(INV 012)。

Example 32

“The phrase is only put it into sentences, so it makes sense. If you do not, so it makes nonsense. So you cannot understand it properly. You have to learn some by heart, then you should learn something that you used most frequently by heart. Others you can read the article or you do some exercise in sentences (INV 001)”.

Example 33

1. “When I learn phatic expressions, I will read them aloud few times. After remembering them, I will use them for the communication to get a deeper impression of those phatic expressions…Since phatic expressions are interesting and funny, you can learn them in a recreational and funny ways (INV 003)”.
   如果记寒暄语的话，我就自己念几遍，这个句子记住之后就会和别人交流，通过交流来记住。因为寒暄语的话它就比较好玩有趣，在记的时候就通过一些有趣的方式呀，在开心中学习。就是那样(INV 003)。
2. “I learned those expressions from movie and books, which is context-dependent, and in this way, I said one can learn through your eyes, also know how to use these expressions (INV 012)”.
   口头禅啊这种的话应该在平时看的电影啊读的书呀，就是你不仅去看还会用(INV 012)。

Example 34

“You can learn a word by word formation. So, you can say that one word, and through the word formation method, you can learn three or four more words. You can change a word, or you can add some elements, as it is changed into another word. So it is very efficient to learn words by words formation (INV 001)”.

Example 35

1. “I am learning vehicle in English, you have to give a text about the vehicle. So I can learn some specific words in that article. Put the phrases and words into sentences of the article, so that it makes sense. Maybe, I don’t know it is right or not. It is just my way (INV 001)”.
2. “I and my classmates were poor-motivated to learn terminologies or compound words, only depended on the teachers’ instruction, it just makes things easier (INV 012)”.
   平时我们讲专业术语比较简单，都是在老师的引导下，才能理解的。如果是自己的话，可能就比较懒(INV 012)。

Example 36

1. “In Chinese, if describe a thing, we need certain methods, but in English, we don't. For example, don't be on my way, if we think in Chinese, it means don't stand on the road, and then transfer that to block someone’s way. But we don't need that kind of transfer in English (INV 001)”.
2. “Based on the previous Chinese expressions, I will find their equivalent phrases, or the synonyms, and then use them (INV 011)”.
   就按照之前的汉语的来想，或者是代替它的词语，或者近义词等。去用这(INV 011)。

Example 37

“For example, he gave two examples, for example, for instance, and then normally, you would not give to many examples, four at most. You can use the reason are as follows, or you can use the word reason, you can replace reason with other words, or, you can use such as, or just use colons, or e.g. or something else (INV 004)”.
举个例子。他说了两个方法 for example, for instance,还有一个就是举例子嘛，因为我觉得不可能举太多，不可能有四五个例子举吧，最多应该有四个。比如说，当举例子的时候，the reasons are as follows, 或者说把 reason，以原因为例，你可以替换 reason，或者还有 such as,或者还有，你可以用冒号，因为冒号也具有举例的作用，还可以用 e.g.或者其他的(INV 004) 。

Example 38

“Hug you into pieces, for example, this phrase is interesting, and in Chinese, there is no such expression. I may remember it easily (INV 011)”.
比如说 hug you into pieces，中文里没有这个表达，很有趣，可能会记住 (INV 011)。

Example 39

“In the high school, the teachers told us it was not good if you used some formulaic language or words again and again in your composition, it would make your audience sleepy; Instead, you have to use formulaic language in different ways and make some changes to that (INV 005)”.
这点的话高中的时候老师就有讲过，就是说同一个单词出现在同一篇文章中很多次不好，什么别人看的都想睡了呀什么的，然后就是说要不一样嘛，要有变化(INV 005)。

Example 40

1. “I just look for the synonyms or change the form of the words (INV 002)”.
   找同义词，或者词形变换(INV 002)。
2. “I just make some changes. I use these expressions earlier, and use the other one later on. Usually, I use some formulaic language with same meaning, but different structure or words in the expression (INV 008)”.
   就是调换一下，前面用这种，后面用另外一种，用相同意思的，不同结构的，或者不同词语的一种表达(INV 008)。

Example 41

“I think, if write for exams, I will use some models of writing, which is already formed in your brain. In this way, you can get higher scores. While, I don’t think it I have to obey

those rules in the exams when I write for dairy. Different from in the composition, in dairy writing, you do need think about the fixed collocations, or fixed sentence types. You can write relative freely. But in composition writing, you cannot do this. You have to think about using some fixed phrase, sentence types and fixed collocations. They are two different kinds of feeling (INV 003)".

我觉得，应付一些考试的话我还是会用一些模版。在大脑中已经形成了，就是说用了模版，拿的分数才会好一点。就会这样。我觉得在周记里面的话就不会受那些条条框框的束缚啊，不会说在我的周记里面我要尽量用这些个固定搭配啊，或者固定句型啊，那时候就不会，就写的比较随意了。但是写作文的话就不能那样了，写作文的话就要想到一些固定的词组啊，句型啊，搭配啊，两种感觉是不同的(INV 003)。

Example 42

1. "Sometimes he changed parts of the formulaic language, but he never changed any word in the famous quotes (INV 004)".
   如果说是纯正的，比如说名言，就不要改(INV 004)
2. "I barely used the memorised long sentences in the English writing', instead, he expressed his ideas in his own words (INV 005)".
   我实际上的话，还没有很长的话去用过的。没有(INV 005)。

Example 43

1. "If you use some formulaic language repeatedly in a text, it looks tedious, and the audience may feel the writer lacks of creative thinking (INV 012)".
   如果这个在一个文章中一直去用的话，我觉得显得就会太累赘了。就会觉得这个文章就会很没有创意(INV 012)。
2. "If you use some phrases all the time, people may think your English level is limited. If you really are good at English, you can use formulaic language in various ways, you can choose other ways of expressions. If you use the same formulaic language in same form, it only showed your limited vocabulary (INV 012)".
   如果你一直反复地用一个词的话就会觉得你这个人的水平吧就不怎么样。如果你真有能力的话，你可以用各种，或者你想表达的话你可以选择其他的方式，如果你一直用这个的话，就显得你的词汇方面就会很有限(INV 012)。

Example 44

"I was worried sometimes, but you do not need over-worried about that. For some formulaic language, you had no way to stop others using it. However, you have to use that in your own ways, for example, you can use different places in your writing, or use in different occasions (INV 012)".

担心是有。但是不用过度的担心。因为一个词块嘛，你不能不让别人用，但是我们用的时候得有自己独特的方式。比方说，你在哪里，或者在那种场合用(INV 012)。

Example 45

"Take any article in the textbook as an example, teachers always gave us simple analysis of the sentences or phrases, and then we can comprehend that by ourselves, mainly outside of the classroom. The analysed parts include some difficult sentences involved

lots grammar stuff, or certain sentences with special structure or some paragraphs (INV 006)".

它基本上如果说一篇课文有什么的话，也就是进行简单的分析，分析了之后也都是自己在下面去理解什么的。一般的分析包括这些还有一些比较难懂的句子呀，比如语法成分比较多的，一些比较难的句子还有段落(INV 006)。

Example 46

"My teacher introduced that they had to learn the writing templates, compound words and terminologies by rote learning, in order to memorise these types of formulaic language accurately and mostly, writing templates would be used in the writing exams, for example, TEM 4; as for the phatic expressions, we could learn that in the pre-scripted dialogues (INV 007)".

像固定文章模板，之前我们考专四的时候或者像原来老师说过像什么三段式啊，就是写议论文啊，还有像短语啊什么的，这就是必须得记得。还有口头禅问候语寒暄语啊这些，老师叫我们背对话。就穿插在文章中。还有就是合成词和专业术语，这些老师就是让我们死记硬背(INV 007)。

Example 47

"Teachers introduced some techniques for memorising formulaic language, such as using association or classification of the items. Also, we were kind of forced to memorise the formulaic language that the teachers taught in the class, since we have to do the dictation on the formulaic language as a means of assessment. You have to do that (INV 003)".

老师让我们记吧，就是可能会有一些什么联想记忆法呀，或者就是分类记忆法呀，或者是没有方法的话直接叫我们记的。就是给我们施加一些硬性压力呀，比如说没有记住的话最终会采取什么样的惩罚措施。我们就会记(INV 003)。

Example 48

1. "Formulaic language teaching in the university is not like that in the school, for instance, like before the university, in the basic language learning stage, the teachers emphasised that we should memorise some kinds of formulaic language, such as fixed expressions or idioms and certain sentence types. However, now, the focus is put on the use formulaic language in a flexible way (INV 009)".

   像大学的话就不像初高中，那种基础阶段，像以前的话，就是注重记，记固定搭配啊固定句式结构啊，像现在的话就是强调的是灵活运用(INV 009)。
2. "When passing exams become the purpose of learning English, teachers would like to pay more attention to the language usage that appears in the exam papers. For example, the conjunctional words in the passage, or some "good beginning sentences" for a composition, and some proverbs, they will think of the benefits that these kinds of language will bring to the writing. Also, readings in the exams gain a lot of attention, for example, how students could comprehend the passages by learning these kinds of language. But once, you entered into the university, nobody asks you to learn this kind of language, and it totally becomes a kind of interest, out of some academic drive (INV 001)".

   当然，当你以 exam 为目的的时候，老师们肯定会更注重高考所需要的这一类的语言的用法。比如说，文章的连接词，再比如说一些好的开头所需要的一些用法，再

比如说催人向上的一些谚语，他们会看重这些在英语写作中的好处。他们在考试中也会很注意考试的阅读，你怎么把它看懂，所以说他们也会提倡这个。但是你到了大学，没有人追着你屁股后面让你学这些东西，完全是出自 Academic 的一种爱好，或者是一种非常学术的，它为什么会这样(INV 001)。

Example 49

"Since it is very hard to learn in books or learn it from teachers, some usage that if you want to speak like the English, so you have to read them in articles, read articles written by them. For example, if you want to learn some beautiful words, you have to read something like Shirley, or something from Shakespeare, if you cannot, so you rarely learn idiomatic usage in English, maybe it is Chinglish (INV 001)".

Example 50

"We can learn some kinds of formulaic language, like terminologies, idioms and fixed expressions in the classroom, whereas some other types of formulaic language are important in speaking, but we have to learn those by themselves outside of the classroom (INV 008)".

我觉得专业术语，俗语和固定搭配在大学的，还有固定文章模版这三个是在大学里面，就是在课堂上能够学到的最多的。然后其他的着重的是口语，就是比较欠缺的。一般就是在家里面看那些美剧啊，自己学(INV 008)。

Example 51

"As writing template, when we prepared for the TEM-4 or earlier in the high school, teachers taught us the three-paragraph essay, we can use that for writing in the exams; also, we were told certain kind of language we have to remember, for instance, phrases. Something like greetings or phatic expressions, teachers asked us to memorise the dialogue in the textbooks. The rest, compound words, and terminologies, we learn through rote learning (INV 007)".

像固定文章模板，之前我们考专四的时候或者像原来高中老师说过像什么三段式啊，就是写议论文啊，还有像短语啊什么的，这就是必须得记得。还有口头禅问候语寒暄语啊这些，老师叫我们背对话。就穿插在文章中。还有就是合成词和专业术语，这些老师就是让我们死记硬背(INV 007)。

Example 52

"There are substantial differences between the formulaic language teaching in the university and at schools, since the university teaching tends to focus on the understanding of the whole text or paragraph, or the underlying meaning that the writer wanted to express; rather than concentrating on specific language use details(INV 010)".

我觉得大学里面跟上高中时候有很大的不同在于，对这些语言，比如俗语和固定搭配啊或者固定文章模板没有那么重视。它不是重视在这些小块上去抓，然后在大的范围，比如对文章整体一个思想方面的理解或者说是(INV 010)。

Example 53

"They could influence the students in an implicit way and make them think in English way, as well as cover the language use in exams and daily use (INV 003)".

就是考试呀还是生活中都是能用到的，就是我觉得这样学的话也可以锻炼我们的一种英文思维嘛(INV 003)。

Example 54

"The content in the textbook is quite limited, and thus probably one cause of the repetition or repeated use of some formulaic language, especially in the English written texts in the class (INV 006)".

我觉得课本上的东西都是比较局限的。课本上的东西你想啊每个人都在学，在你记的基础上，别人也在学(INV 006)。

Example 55

"I think they can meet my learning needs, as they cover most areas that I would use English, especially in terms of formulaic language, especially beneficial for the future professional English language use (INV 009)".

在职场啊，好多专业领域，特定的领域。其他的话，也就是就我们平时的教育这方面(INV 009)。

Example 56

"The content of textbook was limited, and students cannot learn all six types of formulaic language only from the textbooks; for example, the textbook for Advanced English, the interviewee reported that most of the articles they learned in that book are literature works; you cannot find any writing templates we used in the exams from the textbooks; and few formulaic language can be used in current colloquial context or in real life(INV 007)".

觉得不够。比如说我们现在学高英，它编写这个教材，就是涉及到各个方面，像什么文学类啊，还有，主要是文学类的。它就是像什么很多生活化的语言它那个里面就涉及不到，还有像固定文章模板啊在里面也没有反映出来。或者是合成词啥的，比较少(INV 007)。

Example 57

1. "University teachers, I think they sometimes over-emphasize the formulaic language learning in their teaching. For example, the V+ N phrases. I don't think they are so called fixed phrase, or formulaic language. And then, teachers asked us to learn the phrase play table tennis, by heart, as a fixed phrase. I personally feel if it is common in the books, then, it is not a fixed phrase (INV 004)".

   大学老师，也不是说不强调，而是我觉得有时候他们强调过了。比如说，动词短语和名词搭配，我个人认为它就不是短语。比如说，打乒乓球，play table tennis,然后 play 是动词，table tennis 是名词。这个在我看来，它就不算短语。然后，老师就提出来，叫我们记一下，作为短语。我个人认为就是，平常看书就能见到的，根本就不算是什么短语(INV 004)。
2. 'The teachers should give us some more exercise for practice, including some good opening sentences and ending a sentence in the English writing, some good writing samples with comments alongside, some writing templates. I think it is better for improving our ability and levels in English writing (INV 004)".

   老师应该多给我们发一些习题练一练，包括一些作文的开头和结尾，好的模板，好的范文，旁边有点评，提炼的有模板，这样会更好一些。这样会对提升我们的写作能力，和写作水平会更迅速些(INV 004)。

Example 58

1. “Since it is very hard to learn in books or learn it from teachers, some usage that if you want to speak like the English, so you have to read them in articles, read articles written by them. For example, if you want to learn some beautiful words, you have to read something like Shirley, or something from Shakespeare, if you cannot, so you rarely learn idiomatic usage in English, maybe it is Chinglish (INV 001)”.
2. “We can learn some kinds of formulaic language, like terminologies, idioms and fixed expressions in the classroom, whereas some other types of formulaic language are important in speaking, but we have to learn those by themselves outside of the classroom (INV 008)”.
   我觉得专业术语，俗语和固定搭配在大学的，还有固定文章模版这三个是在大学里面，就是在课堂上能够学到的最多的。然后其他的着重的是口语，就是比较欠缺的。一般就是在家里面看那些美剧啊，自己学(INV 008)。

Example 59

“I was thinking, if everyone in the class received the same hand-out of formulaic language from one teacher, I was afraid of that everyone would use the same language, especially in the writing. I then started to think about changing the expression, or create a new way to express myself (INV 006)”.
我就想，如果老师给大家每个人发的都是一样的话，写出来的作文都是差不多的，那就是千篇一律的感觉，那我就想换别的表达方式，或者是来一种创新什么的(INV 006)。

Example 60

“And my teacher Daniel often said that “you are an enthusiasm writer, however your writing is lack of idiomatical convey and what you should do is read more articles written by the native English speakers, and it’s also very important to learn some words by heart and some idiomatical usage learned by heart (INV 001)”.

Example 61

“Once criticised the English Speaking textbook, as the word different used repeatedly across a dialogue and should be changed into other words, for example, differ from; by means of this, the language sounds not that boring (INV 005)”.
我们的口语书里都有，刚举个例子，比如说 different, 我们口语老师就说，说编书的编的不好。他会跟我们说用 differ from 来代替它，就没有那么繁琐也没有那么乏味了(INV 005)。

Example 62

“Through the interaction or communication with the foreign teachers, since they are the native speakers of English, and I can learn the real and authentic English from them. Rather than you learn the language from the textbooks, you can really communicate with the foreigners, both in and outside of the classroom. I learned a lot (INV 012)”.
同他们的交流也会获益很多。比如他们说的，肯定就是，一听就是本土文化，就我们就会迫不及待的想从他们的身上去了解。不是从书本上，而是真正的与外国人接触，而那个时候不管是在课堂上还是在课下与他们打成一片，都会学到不少(INV 012)。

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.

Aijmer, K. (1996). *Conversational routines in English.* London: Longman.

Allen, D. (2010). Lexical bundles in learner writing: An analysis of formulaic language in the ALESS learner corpus. *Komaba Journal of English Education*, 105-127.

Altenberg, B. (1998). On the phraseology of spoken English: the evidence of recurrent word combinations. In A. P. Cowie (Ed.), *Phraseology: theory, analysis and applications* (pp. 101-122). Oxford: Clarendon Press.

Askehave, I., & Swales, J. M. (2001). Genre identification and communicative purpose: A problem and a possible solution . *Applied linguistics, 22(2)*, 195-212.

Bal, B. (2010). *Analysis of Four-word Lexical Bundles in Published Research Articles Written by Turkish Scholars.MA thesis.* Georgia State University.

Ballard, B., & Clanchy, J. (1991). *Teaching Students from Overseas: A Brief Guide for Lecturers and Supervisors.* Melbourne: Longman Cheshire.

Barlow, M. (2005). Computer-based analyses of learner language. In G. Barkhuizen, & R. Ellis (Eds.), *Analysing Learner Language* (pp. 335-357). Oxford: Oxford University Press.

Barnbrook, G. (2009). Sinclair on Collocation. In R. Moon, *Words, Grammar, Text: Revisiting the Work of John Sinclair* (pp. 23-38). Amsterdam: John Benjamins Publishing Company.

Bhatia, V. (2004). *Worlds of written discourse: A genre-based view.* Continuum: London.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings.* London: Longman.

Biber, D. (2006). Stance in spoken and written university registers. *Jornal of English for Academic Purposes, 5*(2), 97-116.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for specific purposes, 26*(3), 263-286.

Biber, D., & Finegan, E. (1994). *Sociolinguistic perspectives on register.* Oxford: Oxford University Press.

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., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English.* London: Longman.

Brown, J. D. (2004). 19 Research Methods for Applied Linguistics: Scope, characteristics, and Standards. In A. Davies, & C. Elder (Eds.), *The Handbook of Applied linguistics.* (pp. 476-500). Blackwell Publishing.

Brown, J., & Rodgers, T. (2002). *Doing Second Language Research.* Oxford: Oxford University Press.

Bryman, A. (2004). *Social Research Methods* (2nd ed.). Oxford: Oxford University Press.

Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of mixed methods research* , 8-22.

Bryman, A. (2012). *Social Research Methods.* Oxford: Oxford University Press.

Bybee, J. (2010). *Language, usage and cognition (Vol.98).* (J. L. Bybee, Ed.) Cambridge: Cambridge University Press.

Bybee, J., & Hopper, P. J. (2001). *Frequency and the emergence of linguistics structure (VOL.45).* (J. L. Bybee, & P. J. Hopper, Eds.) Amsterdam: John Benjamins Publishing.

Byrd, P., & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL, 5*, 31-64.

Caudery, T. (1998). Increasing students' awareness of genre through text transformation exercises: an old classroom activity revisited. *Teaching English as a Second or Foreign Language*, 1-14.

Chamot, A. U. (1987). The learning strategies of ESL students. In A. Wenden, & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 71–84). Englewood Cliffs, NJ: Prentice-Hall.

Chang, H., & Holt, R. (1994). A Chinese perspective on face as inter-relational concern. In S. Ting-Toomey, *The challenge of facework* (pp. 95-132). Albany: State University.

Chen, Y. (2009). *Investigating lexical bundles across learner writing development. PhD thesis.* Lancaster University.

Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology, 14(2)*, 30-49.

Cheng, L. (2008). The key to success: English language testing in China. *Language Testing, 25*(1), 15-37.

CMELT. (2000). *English Curriculum Standards for English Major in China Higher Education.* Shanghai: Foreign Language teaching and research press.

Cobb, T. (2003). Teaching and researching writing. *System, 31*(1), 132-136.

Coffey, A., & Atkinson, P. (1996). *Making sense of qualitative data: complementary research strategies.* Thousand Oaks, CA: Sage.

Cohen, A. (2004). Relevance Theory, Action Theory and Second Language Communication Strategies. *Second Language Research*, 289-302.

Cohen, A. D. (2007). *Language learner strategies thirty years of research and practice.* (A. D. Cohen, Ed.) Oxford: Oxford University Press.

Cohen, A. D. (2011). Second language learner strategies. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 681-698). New York : Routledge.

Cohen, A. D., & Weaver, S. J. (2006). *Styles-and strategies-based instruction: A teachers' guide.* Center for Advanced Research on Language Acquisition.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). London and New York: Routledge.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics, 32*, 45-61.

Cortazzi, M., & Jin, L. (1996). English teaching and learning in China. *Language Teaching, 29*, 61-80.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*(4), 397-423.

Cortes, V. (2008). A Comparative Analysis of Lexical Bundles in Academic History Writing in English and Spanish. *Corpora 3 (1)*, 43-57.

Coulmas, F. (1981). Introduction: conversational routine. In F. Coulmas (Ed.), *Conversational routine* (pp. 1-17). The Hague: Mouton.

Cowie, A. P. (1988). Stable and creative aspects of vocabulary use. In *Vocabulary and language teaching* (pp. 126-139). London: Longman.

Cowie, A. P. (1998). *Phraseology: theory, analysis, and applications.* (A. P. Cowie, Ed.) Oxford: Clarendon Press.

Coxhead, A., & Byrd, P. (2007). Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing, 16*(3), 129-147.

Creswell, J. (1994). *Research design: Qualitative and quantitative approaches.* London: Sage.

Creswell, J. (2003). *Research Design: Qualitative, Quantitative, and Mixed Methods* (2nd ed.). Thousand Oaks, CA: Sage.

Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches.* London: SAGE Publications.

Creswell, J. (2014). *Research Design* (4th ed.). London: Sage.

Creswell, J., & Plano Clark, V. (2011). *Designing and Conducting mixed methods research (2nd).* London: Sage.

Creswell, J., Plano Clark, V., & Garrett, A. L. (2008). Methodological Issues in Conducting Mixed Methods Research Design. In M. M. Bergman (Ed.), *Advances in Mixed Methods Research* (Vol. 28, pp. 66-83). Thousand Oaks, CA: SAGE.

Dai, W. D., & Zhang, X. M. (2005). Tan waiyu zhuanye jiaoxue pinggu he xueke jianshe.[On assessing and establishing English major programme]. *China Foreign Languages, 2*(2), 4-7.

Dai, W. D., & Zhang, X. M. (2007). Dui woguo yingyu zhuanye benke jiaoxue de fansi.[Reflections on the undergraduate English Major progamme]. *Foreign Language World, 4*, 2-11.

De Cock, S. (2004). Preferred sequences of words in NS and NNS speech. *Belgium Journal of English and Literatures (BELL)*, 225-246.

Dechert, H. (1983). How a story is done in a second language. In C. Færch, & G. Kasper (Eds.), *Strategies in Interlanguage Communication* (pp. 175-195). London: Longman.

Denscombe, M. (2008). Communities of practice. *Journal of Mixed Methods Research, 2*, 270-283.

Denzin, N. (1970). *Sociological methods: a source book.* Chicago: Aldine.

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.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies.* Oxford: Oxford University Press.

Doyle, L., Brady, A., & Byrne, G. (2009). An overview of mixed methods research. *Journal of Research in Nursing, 14*(2), 175-185.

Dudley- Evans, T. (1999). The dissertation: A case of neglect? In *Issues in EAP writing Research and Instruction* (pp. 26-38). Reading: Centre for Applied Language Studies, University of Reading.

Dudley-Evans, T., & St John, M. J. (1998). *Developments in English for specific purposes: A multi-disciplinary approach.* Cambridge: Cambridge university press.

Durrant, P., & Mathews-Aydınlı, J. (2011). A function-first approach to identifying formulaic language in academic writing. *English for Specific Purposes, 30*(1), 58-72.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL-International Review of Applied Linguistics in Language Teaching, 47(2)*, 157-177.

Ellis, N. C. (2002). Frequency effects in language acquisition: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*, 143-188.

Ellis, N. C. (2008). Phraseology: The periphery and the heart of language. In S. Granger, & F. Meunier (Eds.), *Phraseology in Foreign Language Learning and Teaching* (pp. 1-13). Amsterdam : John Benjamins.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 17-44.

Ellis, R. (1997). *SLA research and language teaching.* Oxford: Oxford Univeristy Press.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language.* Oxford : Oxford University Press.

Erman, B. (2009). Formulaic language from a learner perspective: What the learner needs to know. In E. A. R. Corrigan (Ed.), *Formulaic language: Volume II: Acquisition, loss, psychological reality, and functional* (pp. 323-346). Amsterdam/Philadelphia: John Benjamins.

Erman, B., & Warren, B. (2000). The idiom principle an the open principle. *Text, 20*(1), 29-62.

Evert, S. (2008). Corpora and collocation. In M. K. Anke Lüdeling, *Corpus linguistics: An international handbook.* (Vol. 1, pp. 223-233). Berlin: De Gruyter Mouton.

Firestone, W. (1987). Meaning in methods: The rhetoric of quantitative and qualitative research. *Educational Researcher, 16*(7), 16-21.

Firth, J. R. (1957). *Studies in linguistic analysis.* Oxford: Blackwell.

Flowerdew, J. (2002a). *Academic Discourse.* London: Pearson Education.

Flowerdew, J. (2002b). Genre in the Classroom: A linguistic approach. In *Genre in the Classroom: Multiple Perspectives* (pp. 91-102). Mahwah, NJ: Lawrence Erlbaum.

Flowerdew, L. (2005). An investigation of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes, 24*(3), 321-332.

Flowerdew, L. (2012). *Corpora and Language Education.* London: Palgrave Macmillan.

Foster, P. (2001). Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers." . In M. Bygate, P. Skehan, & M. Swain, *Researching pedagogic tasks: Second language learning, teaching, and testing* (pp. 75-97). London: Longman.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). Boston: McGraw-Hill.

Galloway, N., & Rose, H. (2015). *Introducing Global Englishes.* London: Routledge.

Gass, S. M., & Selinker, L. (2008). *Second language acquisition: An introductory course.* (3rd ed.). New York : Routledge.

Gay, L. R., Mills, G. E., & Airasian, P. (2006). *Educational Research* (8th ed.). New Jersey: Pearson.

Gilquin, G. (2007). To err is not all. What corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift fur Anglistik und Amerikanistik, 55*, 273–291.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Gorard, S., & Taylor, C. (2004). *Combining methods in educational and social research.* McGraw-Hill International.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocations and formulae. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 145-160). Oxford: Clarendon Press.

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: collocationsand formulae. In *Phraseology: theory, analysis and applications* (pp. 145-160). Oxford: Oxford University Press.

Granger, S. (2009). Learner corpora: A window onto the L2 phrasicon. In A. Barfield, & H.Gyllstad, *Researching collocations in another language* (pp. 60-65). Basingstoke, UK: Palgrave Macmillan.

Granger, S., Dagneaux, E., & Meunier, F. (2002). *The International Corpus of Learner English. Handbook and CD-ROM.* Louvain-la Neuve: Presses universitaires de Louvain.

Greene, J., Caracelli, V. L., & Gramham, W. F. (1989). Towards a Conceptual Framework for Mixed-Method Evaluation Designs. *Educational Evaluation and Policy Analysis, 11*(3), 255-274.

Grotjahn, R. (1987). On the methodological basis of introspective methods. In C. Faerch, & G. Kasper (Eds.), *Introspection in second language research. Multilingual matters, No. 30* (pp. 54-81). Clevedon: England: Multilingual Matters.

Gu, P. Y. (2007). Foreword. In A. D. Cohen (Ed.), *Language learner strategies: thirty years of research and practice* (pp. vii-viii). Oxford: Oxford University Press.

Gui, S., & Yang, H. (n.d.). Retrieved 09 13, 2016, from Chinese Learner English Corpus (CLEC): http://langbank.engl.polyu.edu.hk/corpus/clec.html

Guo, X. (2011). Formulaic sequences: noticing and retrieval by English majors. *Journal of PLA University of Foreign Languages, 34*(2), 44-52.

Hakuta, K. (1976). A case study of a Japanese child learning English as a second language. *Language learning, 26*(2), 321-351.

Halliday, M. A. (1978). *Language as social semiotic.* London: Arnold.

Halliday, M. A. (1985). *An introduction to functional grammar.* London: Edward Arnold.

Halliday, M. A. (1994). *Functions of Language (2nd).* London: Edward Arnold.

Halliday, M. A., & Matthiessen, C. (2004). *Introducing functional grammar.* New York: Edward Arnold.

Han, Z. (2004). *Fossilization in Adult Second Language Acquisition.* Clevedon: Multilingual Matters.

Harklau, L. (2002). The role of writing in classroom second language acquisition. *Journal of Second Language Writing, 11*(4), 329-350.

Harwood, N. (2006). (In)appropriate personal pronoun use in political science: A qualitative study and a proposed heuristic for future research. *Written Communication, 23*, 424-450.

Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics, 4*(2), 237-258.

Horowitz, D. (1986). Essay examination prompts and the teaching of academic writing. *English for Specific Purposes, 5*(2), 107-120.

Hu, G. (2005). 'CLT is best for China'----an untenable absolutist claim. *ELT Journal, 59*(1), 65-68.

Hu, W. Z., & Sun, Y. Z. (2006). On strengthening humanistic education in the English language curriculum. *Foreign language teaching and research, 38*(5), 243-247.

Huang, K. (2014). *Lexical bundles in Chinese EFL learners' essay writing. PhD thesis.* Hongkong: University of Hongkong.

Huang, K. (2015). More does not mean better: Frequency and accuracy analysis of lexical bundles in Chinese EFL learners' essay writing. *System, 53*, 13-23.

Hunston, S. (2002). *Corpora in Applied Linguistics.* Cambridge: Cambridge University Press.

Hüttner, J. (2005). Formulaic language and genre analysis: the case of student academic papers. *VIENNA ENGLISH WORKING PAPERS*, 3-20.

Hüttner, J. (2007). *Academic writing in a foreign language: an extended genre analysis of student texts.* Frankfurt: Peter Lang.

Hüttner, J. (2008). The genre (s) of student writing: developing writing models. *International Journal of Applied Linguistics*, 146-165.

Hyland, K. (2004). *Genre and Second Language Writing.* Ann Arbor: The University of Michigan Press.

Hyland, K. (2008a). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes, 27*(1), 4-21.

Hyland, K. (2008b). Academic clusters: Text patterning in published and postgraduate writing. , 18(1), . *International Journal of Applied Linguistics, 18*(1), 41-62.

Hyland, K. (2009). *Teaching and Researching Writing* (2nd ed.). London: Longman.

Hyland, K. (2011). Academic Discourse. In K. Hyland, & B. Paltridge (Eds.), *The continuum companion to discourse analysis* (pp. 171-184). London: Continuum.

Hyland, K. (2012). Bundles in Academic Discourse. *Annual Review of Applied Linguistics*, 150-169.

Hymes, D. (1984). Sociolinguistics: stability and consolidation. *International Journal of the sociology of Language, 45*, 39-45.

Javis, S., Grant, L., Bikowski, D, & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of second language writing* , 377-403.

Jenkins, J. (2000). *The Phonology of English as an International Language.* Oxford: Oxford University Press.

Jenkins, J. (2007). *English as a lingua franca: attitude and identity.* Oxford: Oxford University Press.

Jin, L., & Cortazzi, M. (2002). English language teaching in China: A bridge to the future. *Asia Pacific Journal of Education, 22*(2), 53-64.

Johnson, D. M. (1992). *Approaches to research in second language learning.* New York : Longman.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research, 1*(2), 112-133.

Juknevičienė, R. (2009). Lexical bundles in learner language: Lithuanian learners vs. native speakers. *Kalbotyra, 61*, 61–72.

Kachru, B. B. (1986). *The alchemy of English.* London: Pergamon Press.

Kachru, Y. (2011). World Englishes. In E. Hinkel (Ed.), *Handbook of research in English language teaching and learning. Volume 2* (pp. 155-172). New York: Routledge.

Kaneyasu, M. (2012). *From Frequency to Formulaicity: Morphemic Bundles and Semi-Fixed Constructions in Japanese Spoken Discourse. PhD thesis.* Los Angeles: University of California.

Kelty, J. (1991). *Family AlbuM, U.S.A.* New York: MacMillan.

Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching.* Cambridge: Cambridge University Press.

Kirkpatrick, R., & Zang, Y. (2011). The negative influences of exam-oriented education on Chinese high school students: backwash from classroom to child. *Language Testing in China*, 36-45.

Kuiper, K., & Haggo, D. (1984). Livestock auctions, oral poetry, and ordinary language. *Language in society, 13*, 205-234.

Lamb, S. M. (1998). *Pathways of the brain: The neurocognitive basis of language.* Amsterdam: John Benjamins Publishing.

Langacker, R. (1986). An introduction to cognitive grammar. *Cognitive Science, 10*, 1-40.

Larsen-Freeman, D., & Long, M. (1991). *An introduction to second language acquisition research.* Harlow: Longman.

Lee, D., & Chen, S. (2009). Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing, 18*, 181-196.

Leedham, M. (2011). *A corpus-driven study of features of Chinese students' undergraduate writing in UK universities, PhD thesis.* London: The Open University.

Li, H., Meng, L., Shi, X., & Wu, B. (2012). Does attending elite colleges pay in China? *Journal of Comparative Economics, 40*, 78-88.

Li, J., & Schmitt, N. (2009). The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing, 18*(2), 85-102.

Liao, X. (2004). The need for communicative language teaching in China. *ELT journal, 58*(3), 270-273.

Lichtman, M. (2012). *Qualitative research in education: A user's guide* (3rd ed.). Thousand oaks: California: Sage.

Lillis, T. (2008). Ethnography as methods, methodology and "deep theorizing.". *Written Communication*(25), 353-388.

Liu, W., & Huo, Y. (2011). On the role of formulaic sequences in second language acquisition. *US-China Foreign Language*, 708-713.

Lorenz, G. (1999). *Adjective Intensification – Learners Versus Native Speakers: A Corpus Study of Argumentative Writing.* Amsterdam: Rodopi.

MacKay, S. (2012). 8.English as an International Lingua Franca Pedagogy. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 122-139). London: Routledge.

Mackay, S., & Rubdy, R. (2009). The social and Sociolinguisitic Context of Language Learning and Teaching. In M. H. Long, & D. C. J, *The handbook of Language Teaching* (pp. 1980-1996). Chichester: Wiley-Blackwell.

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design.* London: Routledge.

Mackey, A., and Gass,S.M. (2005). *Second Language Research: Methodology and Design.* Mahwah, N.J.: Lawrence Erlbaum.

Manchón, R. (2009). *Writing in foreign language contexts: Learning, teaching, and research (Vol. 43).* Bristol: Multilingual Matters.

Manchón, R. (2011). *Learning-to-write and Writing-to-learn in an Additional Language (Vol. 31).* Amsterdam: John Benjamins Publishing.

Martin, J. R. (2009). Genre and language learning: A social semiotic perspective. *Linguistics and Education, 20*(1), 10-21.

Mauranen, A. (2007). *English as a lingua franca: speakers not learners.* Taiwan: University of Helsinki.

Mauranen, A. (2012). *Exploring ELF: Academic English shaped by non-native speakers.* Cambridge: Cambridge University Press.

Maxwell, J. A. (1992). Understanding and validity in qualitative research. *Havard Educational Review, 62(3)*, 279-300.

Mayor, B. (2006). Dialogic and hortatory features in the writing of Chinese candidates for the IELTS test. *Language, Culture & Curriculum, 19*(1), 104-121.

McCrone, J. (1999). States of mind. *New Scientist*, 30-31.

McDonough, j., & McDonough, S. (1997). *Research Methods for English language teachers.* London: Arnold.

Merriam, S. (2009). *Qualitative research: A guide to design and implementation.* San Francisco: Jossy-Bass.

Ministry of Education. (2005). *Guanyu jinyibu jiaqiang gaodeng yuanxiao benke jiaoxue gongzuo tigao jiaoxue zhiliang de ruogan yijian[Guidelines for further strengthening undergraduate education and improving the quality of undergraduate programs]*. Retrieved 11 4, 2013, from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_734/200507/8296.html

Ministry of Education. (2013). *List of Regular Institutions of Higher Education in China.* Retrieved 07 25, 2013, from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_2812/200906/48836.html

Mitchell, R., & Martin, C. (1997). Rote learning, creativity and'understanding'in classroom foreign language teaching. *Language Teaching Research, 1*(1), 1-27.

Modiano, M. (1999). Standard English(es) and educational practices for the world's lingua franca. *English Today, 15(4)*, 3-13.

Moon, R. (1992). Textual aspects of fixed expression in learner's dictionaries. In P. Arnaud, & B. H (Eds.), *Vocabulary and Applied Linguistics* (pp. 13-27). Basingstoke: Palgrave Macmillan.

Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach.* Oxford: Oxford University Press.

Moore, T., & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes, 4*(1), 43-66.

Moran-Ellis, J., Alexander, V. D., Cronin, A., Dickinson, M., Fielding, J., Sleney, J., & Thomas, H. (2006). Triangulation and integration: processes, claims and implications. *Qualitative Research, 6*(1), 45-59.

Morgan, D. L. (1998). Practical strategies for combining qualitative and quantitative methods: Applications to health research. *Qualitative Health Research, 8*(3), 362-376.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained methodological implications of combining qualitative and quantitative methods. *Journal of mixed methods research, 1*(1), 48-76.

Myles, F., & Cordier, C. (2016). The psycholinguistic construct of formulaicity in second language learners: issues of conceptualisation and identification. *FLaRN (Formulaic Language Research Network) 2016 workshop.* Swansea: Swansea University.

Myles, F., & Cordier, C. (2016). The Psycholinguistic construct of formulaicity in second language learners: issues of conceptualisation and identification. *Formulaic Research Network Workshop.* Swansea: Swansea University.

Myles, F., Hooper, J., & Mitchell, R. (1998). Rote or Rule? Exploring the Role of Formulaic Language in Classroom ForeignLanguage Learning. *Language Learning, 48*(3), 323-364.

Nation, P. (1995). (Interviewed by N. Schmitt) The word on words: an interview with Paul Nation. *Language Teacher*, 5-7.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching.* Oxford: Oxford University Press.

Nesslhauf, N. (2005). *Collocations in a learner corpus.* Amsterdam: John Benjamins.

Newman, I., & Benz, C. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum.* Carbondale: Southern Illinois Univerisity Press.

Nunan, D. (1992). *Research methods in language learning.* Cambridge: Cambridge University Press.

Nunan, D., & Bailey, K. (2009). *Exploring Second Language Classroom Research: A Comprehensive Guide.* Boston: Heinle Cengage Learning.

O'Donnell, M. B., Römer, U., & Ellis, N. C. (2015). The development of formulaic language in first and second language writing: Investigating effects of frequency, association, and native norm. In S. Hoffmann, B. Fischer-Starcke, & A. San, *Current issues in Phraseology* (pp. 83-108). Amsterdam: John Benjamins.

Oakey, D. (2009). Fixed collocational patterns in isolexical and isotextual versions of a corpus. *Contemporary corpus linguistics*, 140-159.

O'keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge: Cambridge University Press.

Oxford, R. L. (1990). *Language learning strategies: What every teacher should know.* New York: Newbury House.

Paikeday, T. M. (1985). *The native speaker is dead.* Toronto: Paikeday Publishing Inc.

Paltridge, B. (2001). *Genre and the language learning classroom.* Ann Arbor, MI: University of Michigan Press.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguisitics*, 130-149.

Patton, M. Q. (1980). *Qualitative Evaluation and Research Methods.* London: Sage.

Patton, M. Q. (1990). *Qualitative evaluation and research methods.* Thousand Oaks, CA: Sage Publications.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: nativelike selection and native fluency. In J. Richards, & S. Schmidt (Eds.), *Language and Communication* (pp. 191-226). London: Longman.

Pecorari, D. (2009). Formulaic Language in Biology: A Topic-specific Investigation. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 91-106). London: Continuum.

Peters, A. (1983). *Units of language acquisition.* Cambridge: Cambridge University Press.

Pinker, S. (1999). *Words and rules: The ingredients of language.* New York: Harper-Collins.

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Science*, 456-463.

Polio, C. (2012). Editor's Introduction. *Annual Review of Applied Linguistics, 32*, vi-vii.

Polkinghorne, D. E. (2005). 'Language and meaning: Data collection in qualitative research'. *Journal of Counseling Psychology, 52*(2), 137-145.

Pollastek, A., Hyona, J., & Bertram, R. (2000). The role of morphological constituents in reading Finish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, 820-833.

Punch, K. F. (2005). The analysis of qualitative data. In K. F. Punch (Ed.), *Introduction to Social Research, Quantitative and Qualitative Approaches* (2nd ed., pp. 197-216). London: Sage.

Rampton, M. (1990). Displacing the "native speaker": expertise, affiliation and inheritance. *ELT Journal 44(2)*, 338-343.

Rao, Z. (2002). Chinese students' perceptions of communicative and non-communicative activities in the EFL classroom. *System, 30*(1), 85-105.

Rao, Z. (2013). Teaching English as a foreign language in China: Looking back and forward. *English Today, 29*(3), 34-39.

Reichardt, C. S., & Cook, T. D. (1979). *Qualitative and quantitative methods in evaluation research* (Vol. 1). Beverly Hills, CA: Sage.

Reppen, R. (1995). A genre-based approach to content writing instruction. *TESOL Journal, 4*(2), 32-35.

Reppen, R. (2009). Exploring L1 and L2 writing development through collocations: A corpus-based look. In A. Barfield, & H. Gyllstad, *Researching collocations in another language* (pp. 49–59). Basingstoke, UK: Palgrave Macmillan.

Reppen, R. (2010). *Using corpora in the language classroom.* Cambridge: Cambridge University Press.

Richards, J., Platt, J., & Platt, H. (1992). *Longman Dictionary of Language Teaching and Applied Linguistics.* Harlow: Longman.

Richards, K. (2003). *Qualitative inquiry in TESOL.* London: Palgrave MacMillan.

Robinson, P., & N. Ellis, N. (2008). *Handbook of Cognitive Linguistics and Second Language Acquisition.* London: Routledge.

Rubin, J. (2013). Teaching Language-Learning Strategies. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics* (pp. 1-5). Oxford: Wiley-Blackwell.

Ruth, B. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics, 18*(2), 229-253.

Salazar, D. J. (2011). *Lexical bundles in scientific English: A corpus-based study of native and non-native writing, PhD Thesis.* Barcelona: Universitat de Barcelona.

Saragni, S., & Candlin, C. (2003). Introduction Trading between reflexivity and relevance: new challenges for applied linguistics. *Applied Linguistics, 24*(3), 271-285.

Schmitt, N. (2000). *Vocabulary in Language Teaching.* Cambridge: Cambridge University Press.

Schmitt, N. (2004). *Formulaic Sequences: Acquisition, Processing, and Use.* Amsterdam: John Benjamins.

Schmitt, N., & Carter, R. (2004). Formulaic sequences in action. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 1-22). Amsterdam: John Benjamins Publishing.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid. In N. Schmitt (Ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 127-51). Amsterdam: John Benjamins.

Schwandt, T. (2000). ''Three Epistemological Stances for Qualitative Inquiry: Interpretivism, Hermeneutics and Social Constructionism''. In N. Denzin, & L. Y.S. (Eds.), *Handbook of Qualitative Research* (pp. 189-214). Thousand Oaks, California: Sage Publications.

Scott, M. (2012). *WordSmith Tools version 6.0.* Stroud: Lexical Analysis Software.

Seidlhofer, B. (2006). English as a Lingua Franca in the Expanding Circle: What it Isn't. *English in the world: Global rules, global roles*, 40.

Seliger, H. W., & Shohamy, E. (1989). *Second language research methods.* Oxford: Oxford University Press.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching, 10*, 209-232.

Simpson, R. (2004). Stylistic features of academic speech: The role of formulaic expressions. In U. Connor, & T. A. Upton (Eds.), *Discourse in the professions: Perspectives from corpus linguistics* (pp. 37-64). Amsterdam: John Benjamins Publishing.

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List (AFL). *Applied Linguistics, 31*, 487-512.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Sinclair, J. M. (2004). *Trust the text.* London: Routledge.

Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, 285-301.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Stebbins, R. A. (2001). *Exploratory research in the social sciences.* London: Sage.

Stern, H. H. (1983). *Fundamental concepts of language teaching.* Oxford: Oxford University Press.

Stevick, E. (1989). *Success with Foreign Languages: Seven Who Achieved It and What Worked for Them.* New York: Prentice-Hall.

Stubbs, M. (1996). *Text and Corpus Analysis.* Oxford: Blackwell.

Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics.* Oxford: Blackwell.

Stubbs, M., & Barth, I. (2003). Using recurrent phrases as text-type. *Functions of language 10(1)*, 61-104.

Suter, W. N. (2012). *Introduction to Educational Research* (2nd ed.). London: Sage.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge: Cambridge Univeristy Press.

Swales, J. M. (2004). *Research Genres: Explorations and Applications.* Cambridge: Cambridge University Press.

Swales, J. M. (2008). Foreword. In D. Belcher, & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other interactions* (pp. v-viii). Ann Arbor: University of Michigan Press.

Talmy, S., & Richards, K. (2011). Theorizing qualitative research interviews in applied linguistics. *Applied Linguistics, 32*(1), 1-5.

Tashakkori, A., & Creswell, J. W. (2007). Editorial: The new era of mixed methods. *Journal of mixed methods research, 1*(1), 3-7.

Tashakkori, A., & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches.* Thousand Oaks, CA: Sage.

Tashakkori, A., & Teddlie, C. (2003). *Handbook of Mixed Methods in Social and Behavioral Research.* Thousand Oaks: Sage.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research.* Thousand Oaks, CA: SAGE.

(2011). *The impact of a university pre-sessional course on the academic writing behaviours of a group of Chinese undergraduate students studying for a degree in media and cultural studies.* Lancaster University.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Cambridge, MA: Harvard University Press.

Tomlinson, B. (2005). English as a foreign language: Matching procedures to the context of learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning* (pp. 137-153). Marwah, NJ: Lawrence Erlbaum Associates.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning 61*, 569–613.

Turner, J. (2012). Academic Literacies: Providing a Space for the Socio-political Dynamics of EAP. *Journal of English for Academic Purposes, 11(1)*, 17-25.

Waibel, B. (2008). *Phrasal verbs: German and Italian learners of English compared.* Saarbrucken, Germany: VDM.

Wang, J., & Cortazzi, M. (2013). Changing cultural ways with praise: a distant action research project in China. In M. Cortazzi, & J. L, *Research cultures of learning: international perspectives on language learning and education* (pp. 95-115). London: Palgrave Macmillan.

Warren, B. (2005). A model of idiomaticity. *Nordic Journal of English Studies, 4(1)*, 35-54.

Weinert, R. (1995). The role of formulaic language in second language acquisition: a review. *Applied Linguistics, 16*(2), 180-205.

Weinert, R. (2010). Formulaicity and usage-based language: Linguistic, psycholinguistic and acquisitional manifestations. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 1-20). London: Continuum.

Wellington, J., & Szczerbinski., M. (2007). *Research methods for the social sciences.* London: Bloomsbury Publishing.

Wen, Q. F., Ding, Y., & Wang, W. Y. (n.d.). Features of oral styles in English compositions of advanced Chinese EFL learners. *Foreign Language Teaching and Research, 4*, 268-274.

Williams, J. (2012). The potential role (s) of writing in second language development. *Journal of Second Language Writing, 21*(4), 321-331.

Wood, D. (2010). *Perspectives on Formulaic Language: acquisition and communication.* (D. Wood, Ed.) London: Continuum International Publishing Group.

Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction.* London: Bloomsbury Publishing.

Wray, A. (1992). *The focusing hypothesis: The theory of left hemisphere lateralised language re-examined (Vol. 3).* Amsterdam: John Benjamins.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language & communication 18(1)*, 47-67.

Wray, A. (2002). *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.

Wray, A. (2005). Making sense of patterns: psycholinguistic perspectives on corpus linguistic research. Birmingham University: Plenary paper given at the Corpus Linguistics Conference.

Wray, A. (2008). *Formulaic language: Pushing the Boundaries.* Oxford: Oxford University Press.

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, 231-254.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language & Communication, 20(1)*, 1-28.

Yang, J. (2006). Learners and users of English in China. *English Today, 22*(2), 3-9.

Yu, L. (2001). Communicative language teaching in China: Progress and resistance . *TESOL quarterly, 35*(1), 194-198.

Yu, X. (2015). From Dual-nature View of Language to Audiolingualism: A Reappraisal of Memory Aspect of Language. *Theory and Practice in Language Studies, 5*(3), 493.

Zhong, M. S. (2006). Views on deepening the reform for English language education. *Foreign language teaching and research, 5*(38), 254-256.

Zhu, H. (2003). Globalization and new ELT challenges in China. *English Today, 19*(4), 36-41.

Zou, S. (2003). The alignment of language teaching syllabus and language testing: TEM-8 test development and administration. *Foreign Language World, 98*(6), 71-78.