



外语教学指导与学术研究系列丛书

# Acquisition of English Verbal Formulaic Sequences by Chinese EFL Intermediate Classroom Learners

刘宁 著

 北京理工大学出版社  
BEIJING INSTITUTE OF TECHNOLOGY PRESS

外语教学指导与学术研究系列丛书

**Acquisition of English Verbal  
Formulaic Sequences  
by Chinese EFL Intermediate  
Classroom Learners**

刘 宁 著



**北京理工大学出版社**

BEIJING INSTITUTE OF TECHNOLOGY PRESS

版权专有 侵权必究

---

### 图书在版编目 (CIP) 数据

中国学习者对于英语动词习惯性搭配的习得 = Acquisition of English Verbal Formulaic Sequences by Chinese EFL Intermediate Classroom Learners; 英文/刘宁著. —北京: 北京理工大学出版社, 2012. 7

ISBN 978 - 7 - 5640 - 6261 - 3

I. ①中… II. ①刘… III. ①英语 - 动词 - 研究 IV. ①H314. 2

中国版本图书馆 CIP 数据核字 (2012) 第 153242 号

---

---

出版发行 / 北京理工大学出版社

社 址 / 北京市海淀区中关村南大街 5 号

邮 编 / 100081

电 话 / (010)68914775(办公室) 68944990(批销中心) 68911084(读者服务部)

网 址 / <http://www.bitpress.com.cn>

经 销 / 全国各地新华书店

印 刷 / 保定市中国画美凯印刷有限公司

开 本 / 710 毫米 × 1000 毫米 1/16

印 张 / 10.75

字 数 / 175 千字

责任编辑 / 梁铜华

版 次 / 2012 年 7 月第 1 版 2012 年 7 月第 1 次印刷

责任校对 / 周瑞红

定 价 / 32.00 元

责任印制 / 王美丽

---

图书出现印装质量问题, 本社负责调换

# **Abstract**

**Author: Liu Ning**

**Date: 26 April 2012**

**Title: Acquisition of English Verbal Formulaic Sequences by Chinese EFL Intermediate Classroom Learners**

This study investigates the acquisition of English verbal formulaic sequences (VFS) by Chinese EFL (English as a Foreign Language) intermediate college students in an attempt to find out the major factors influencing their acquisition of these sequences. 70 participants were selected from the four-year non-English-major college students and the three-year English-major college students who were admitted to Beijing Institute of Technology (BIT) in 2007. Among the 70 participants, 43 are four-year non-English-major students who are taken as the higher intermediate group and 27 are three-year English-major students who are taken as the lower intermediate group.

Quantitative data were collected through the participants' English scores in the National Matriculation English Test (NMET), the Vocabulary-size Pre-test and Post-test, the English Verbal Formulaic Sequences Pre-test, Training Test and Post-test, as well as the participants' essays, written and collected at an interval of two or three weeks for a period of half of the participants' first academic year (about four months). Qualitative data were obtained through the questionnaire and the classroom observations.

My data analysis indicates that the major factors hindering the acquisition of English verbal formulaic sequences by the Chinese EFL intermediate learners are the opaque nature of some of the sequences, lack of genuine communicative needs, and inadequate language input, among which the semantic opaqueness of some of the sequences is by far the biggest obstacle. English verbal formulaic sequences with transparent meaning are acquired earlier and better than those with opaque meaning. Furthermore, due to these obstacles in acquiring English verbal formulaic sequences, the participants' mastery of these sequences lags behind other linguistic aspects such as general vocabulary knowledge. The study also finds that the

mastery of English verbal formulaic sequences is closely connected with vocabulary size, general vocabulary knowledge and overall language proficiency.

Based on the findings of this study, some suggestions were put forward for Chinese EFL classroom learning and teaching of English verbal formulaic sequences. Firstly, Chinese EFL learners as well as instructors need pay more attention to the acquisition of English verbal formulaic sequences because the use of L2 formulaic language can facilitate language processing, enhance fluency, and make the learners' utterances more native-like. Secondly, the teaching material should be so compiled as to containing more formulaic sequences and be sequenced according to the semantic transparency/opaqueness of formulaic sequences, with the most transparent ones coming first and the most opaque ones being left to a later stage. Thirdly, classroom instructors should create an atmosphere which favours the learning of formulaic sequences and provide learners with sufficient chances to encounter and practice these sequences. Fourthly, in helping learners facilitate the process of formulaic sequence acquisition, the instructors should draw special attention from the participants to the importance of formulaic sequences and increase the learners' overall English proficiency, general vocabulary and vocabulary size, because these factors are found to be closely connected with the acquisition of English verbal formulaic sequences.

**Key Words:** Verbal Formulaic Sequences, Semantic Transparency, Input, Vocabulary Size, General Vocabulary Knowledge, Overall Language Proficiency

## **Acknowledgements**

Many people and organisations have provided assistance, encouragement and advice during the preparation of this thesis and without them, and many others too numerous to name, this thesis could not have been completed.

First and foremost, my great thanks will be extended to my supervisors Dr. Michael Daller (Director of Study), Dr. Jeanine Caroline Treffers-Daller, and Dr. Jeanette Sakel, who have provided me with valuable instructions and advice. My grateful thanks also go to Dr. Kate Beeching and many other teachers and staff members in HLSS of UWE who have offered help and support to me.

Secondly, my sincere gratitude goes to the authority of BIT (Beijing Institute of Technology) who has offered financial support for my tuition as well as many other conveniences for the completion of this thesis.

Thirdly, I am also grateful to my colleagues, home and international, as well as all the student participants in BIT who willingly invest their time and energy in the data collection of this thesis.

Finally, my great thanks also go to all my family members, relatives and friends who have given me great support and help all the way through the whole process of the completion of this thesis.

## **Key to Abbreviations and Acronyms**

L1	First (Native) Language
L2	Second Language
L3	Third Language
L4	Fourth Language
EFL	English as Foreign Language
BIT	Beijing Institute of Technology
VFS	Verbal Formulaic Sequences
NMET	National Matriculation English Test
SLA	Second Language Acquisition
CECR	College English Curriculum Requirements
ESL	English as Second Language
FL	Foreign Language
TL	Target Language
BNC	British National Corpus
NST	Native-speaker English Teacher
NST1	Native-speaker English Teacher No. 1
NST2	Native-speaker English Teacher No. 2
NNST	Non-native-speaker English Teacher
NNST1	Non-native-speaker English Teacher No. 1
NNST2	Non-native-speaker English Teacher No. 2
NNSS	Non-native-English-speaker Student
NNSS1	Non-native-English-speaker Student No. 1
NNSS2	Non-native-English-speaker Student No. 2
NDW	Number of Different Words
TTR	Type/Token Ratio
LFP	Lexical Frequency Profile

## List of Tables and Figures

### Tables

2.1	Different Types of Formulaic Sequences According to Semantic Transparency .....	23
2.2	Categories of Lexical Phrases and Their Fixedness .....	29
2.3	Components of Word Knowledge .....	41
2.4	What Is Involved in Knowing a Word .....	42
2.5	Facets of Validity .....	45
3.1	Overview of the Number of Participants in the Present Study .....	61
3.2	Summary of the Participants' NMET Score Percentages ( $n=100$ ) .....	61
3.3	Number of English Verbal Formulaic Sequences for Each Transparency/Opaqueness Rank .....	68
3.4	Summary of Topics for the Participants' Essays .....	69
4.1	Group Statistics for the Vocabulary-size Pre-test ( $n=18$ ) .....	77
4.2	Group Statistics for the Vocabulary-size Post-test ( $n=18$ ) .....	77
4.3	Paired Samples Statistics for Vocabulary-size Pre-test and Post-test ( $n=18$ ) .....	77
4.4	Group Statistics for the English Verbal Formulaic Sequences Pre-test ( $n=100$ ) .....	79
4.5	Group Statistics for the English Verbal Formulaic Sequences Post-test ( $n=100$ ) .....	80
4.6	Group Statistics for the English Verbal Formulaic Sequences Training Test ( $n=100$ ) .....	80
4.7	Paired Samples Statistics for the English Verbal Formulaic Sequences Pre-test and Post-test ( $n=100$ ) .....	81
4.8	Summary of the Percentage of Correct Answers for the Items in Each Rank in the English Verbal Formulaic Sequences Pre-test .....	84
4.9	Non-parametric K Related Friedman Test for Each of the Five Ranks on the Transparency/Opaqueness Scale of the English Verbal Formulaic Sequences Pre-test .....	84





4.10	Summary of the Percentage of Correct Answers for the Items in Each Rank in the English Verbal Formulaic Sequences Post-test.....	85
4.11	Non-parametric K Related Friedman samples Test for Each of the Five Ranks on the Transparency/Opaqueness Scale of the English Verbal Formulaic Sequences Post-test.....	86
4.12	Summary of the Percentage of Correct Answers for the Items in Each Rank in the English Verbal Formulaic Sequences Training Test.....	87
4.13	Non-parametric K Related Friedman Test for Each of the Five Ranks on the Transparency/Opaqueness Scale of the VFS Training Test.....	88
4.14	Summary of the Mean Scores and Mean Ranks of the Five Groups of English Verbal Formulaic Sequences in the Three English Verbal Formulaic Sequences Tests .....	88

## Figures

2.1	Model of Speech Production .....	33
2.2	The Lexical Space: Dimensions of Word Knowledge and Ability .....	44

# CONTENTS

<b>Introduction .....</b>	<b>1</b>
<b>Chapter One: Acquisition of Formulaic Sequences .....</b>	<b>11</b>
1.1 First Language (L1) Acquisition of Formulaic Sequences .....	11
1.2 Second Language (L2) Acquisition of Formulaic Sequences .....	14
<b>Chapter Two: The Nature of Formulaic Sequences and                     How to Learn and Teach Them .....</b>	<b>19</b>
2.1 The Holistic and Analytic Features of Formulaic Sequences .....	19
2.2 The Transparency and Opaqueness of Formulaic Sequences .....	21
2.3 The Fixedness of Formulaic Sequences .....	24
2.4 The Processing of Formulaic Sequences .....	31
2.5 The Input of Formulaic Sequences for L2 Learners .....	35
2.6 Vocabulary Knowledge and Formulaic Sequences .....	37
2.6.1 What Is a Word? .....	37
2.6.2 What Does It Mean to Know a Word? .....	39
2.6.3 Measurement of Vocabulary Knowledge .....	44
2.6.4 Vocabulary Knowledge and Formulaic Sequences .....	50
2.7 Structural Differences Between English and Chinese Verbal Formulaic Sequences .....	50
2.8 The Effect of Classroom Instruction on the Acquisition of Formulaic Sequences .....	52
2.9 Overall Language Proficiency and Formulaic Sequences .....	55
2.10 Individual Differences and the Acquisition of Formulaic Sequences .....	56
<b>Chapter Three: Study Questions and Methodology .....</b>	<b>58</b>
3.1 Hypotheses for the Present Study and Their Operationalisations .....	58
3.2 Methodology .....	60
3.2.1 Participants .....	60
3.2.2 Instruments and Procedures .....	62



3.2.2.1	Identifying the Target Verbal Formulaic Sequences .....	62
3.2.2.2	The English Verbal Formulaic Sequences Pre-test, Post-test and Training Test .....	64
3.2.2.3	The Vocabulary-size Pre-test and Post-test .....	65
3.2.2.4	The Questionnaire .....	66
3.2.2.5	Survey of the Semantic Transparency/Opaqueness of the 100 Target English Verbal Formulaic Sequences .....	66
3.2.2.6	The Class Observations .....	68
3.2.2.7	The Participants' Essays .....	68
3.2.2.8	Survey of the Exposure to the 100 Target English Verbal Formulaic Sequences .....	70
<b>Chapter Four: Data Analyses .....</b>		<b>71</b>
4.1	Data Analyses for the Questionnaire .....	71
4.1.1	The Participants' Knowledge of English Verbal Formulaic Sequences .....	71
4.1.2	The Participants' Report on Their Use of English Verbal Formulaic Sequences .....	72
4.1.3	The Input of English Verbal Formulaic Sequences for the Participants .....	73
4.1.4	The Output of English Verbal Formulaic Sequences of the Participants .....	74
4.1.5	Summary .....	75
4.2	Group Differences in Overall Proficiency, Vocabulary Size and the Mastery of English Verbal Formulaic Sequences .....	76
4.2.1	Group Differences in Overall English Language Proficiency .....	76
4.2.2	Group Differences in Vocabulary Size .....	76
4.2.2.1	The Vocabulary-size Pre-test .....	76
4.2.2.2	The Vocabulary-size Post-test .....	77
4.2.2.3	The Vocabulary Size Increase .....	77
4.2.3	Group Differences in the Mastery of English Verbal Formulaic Sequences .....	78
4.2.3.1	The English Verbal Formulaic Sequences Pre-test .....	79
4.2.3.2	The English Verbal Formulaic Sequences Post-test .....	79
4.2.3.3	The English Verbal Formulaic Sequences Training Test .....	80

4.2.3.4 Increase in the Mastery of English Verbal Formulaic Sequences.....	80
4.3 Relationship Between Vocabulary Size and the Mastery of English Verbal Formulaic Sequences .....	81
4.4 Relationship Between Overall English Proficiency and the Mastery of English Verbal Formulaic Sequences.....	82
4.5 Relationship Between the Semantic Transparency and the Mastery of English Verbal Formulaic Sequences.....	83
4.6 Relationship Between Input Frequency and the Mastery of English Verbal Formulaic Sequences.....	90
4.6.1 Textbook Input Frequency and the Mastery of English Verbal Formulaic Sequences .....	90
4.6.2 Explicit Classroom Instructions and the Mastery of English Verbal Formulaic Sequences.....	91
4.7 Relationship Between the Participants' General Vocabulary Knowledge and the Mastery of English Verbal Formulaic Sequences.....	93
<b>Chapter Five: Findings and Discussion .....</b>	<b>95</b>
5.1 Major Findings.....	95
5.2 Discussion.....	100
5.3 Limitations.....	101
<b>APPENDICES .....</b>	<b>102</b>
Appendix I: The 50 Most Frequent Verbs of the English Language .....	102
Appendix II: The Top-10 Verbs with Their Instances in College English Curriculum Requirements (CECR).....	104
Appendix III: The 100 Target English Verbal Formulaic Sequences .....	105
Appendix IV: The Final 100 Target English Verbal Formulaic Sequences in Sentence-length Contexts .....	109
Appendix V: The English Verbal Formulaic Sequences Pre-test and Post-test .....	114
Appendix VI: The English Verbal Formulaic Sequences Training Test .....	119
Appendix VII: The Vocabulary-size Pre-test and Post-test .....	127
Appendix VIII: The Questionnaire.....	129



# Acquisition of English Verbal Formulaic Sequences

by Chinese EFL Intermediate Classroom Learners

---

Appendix IX: The Transparency/Opaqueness Survey of the 100 Target English Verbal Formulaic Sequences .....	132
Appendix X: The Textbook Input Frequency of the 100 Target English Verbal Formulaic Sequences .....	137
<b>References .....</b>	<b>141</b>

# Introduction

Many terms have been used to describe formulaic language. Wray (2002) summarises more than fifty terms, including “amalgams”, “clichés”, “co-ordinate constructions”, “complex lexemes”, “composites”, “fossilised forms”, “frozen metaphors”, “frozen phrases”, “gambits”, “gestalt”, “prefabricated chunks”, “fixed expressions”, “collocations”, “word combinations”, “formulas”, “ready-made expressions”, “unanalysed expressions”, “idioms”, “conventionalised expressions”, “multi-word units”, “institutionalised expressions”, “unanalysed wholes”, etc. (Wray 2002: 9). Albeit many different terms are used by researchers, the research area is almost the same: conventionalised fixed expressions. “Criteria used to identify formulaic language vary according to the focus of research...but there is considerable overlap across studies” (Weinert 1995: 199).

The various terms for the study of formulaic sequences are the result of different criteria used by researchers for defining a formulaic sequence. For example, Myles, Mitchell and Hooper (1999) suggest six criteria for the identification of formulaic sequences on the basis of the criteria offered by Peters (1983, 1985) and Weinert (1995). The first criterion is the greater length and complexity of formulaic sequences. According to this criterion, there should be at least two morphemes in length in a formulaic sequence and the sequence has greater complexity compared with the learner’s other output. The second criterion is phonological coherence. A formulaic sequence must be phonologically coherent, that is, it can be fluently articulated without any hesitation. The third criterion concerns the appropriateness of formulaic sequences. A formulaic sequence may well be inappropriate syntactically, semantically or pragmatically or otherwise idiosyncratic. The fourth criterion is the fixedness of formulaic sequences. A formulaic sequence is used repeatedly and always in the same form with limited or no substitutability of its constituent parts. The fifth criterion relates to the grammaticality of formulaic sequences. A formulaic sequence is unrelated to the learner’s productive patterns, that is, the learner’s formulaic language is well formed and grammatically advanced in comparison with the rest of his/her inter-language. The last criterion is about the context dependence of formulaic

sequences. A formulaic sequence tends to be situationally dependent, that is, it is used in situationally specific ways and predictable in context (Myles et al. 1999: 51-52).

Wray (2002) suggests that formulaic sequences have a number of characteristics in common in form, function, meaning and provenance, which may help us to locate them. As to the form of formulaic sequences, three features should be borne in mind: irregularity, variability and collocation. Irregularity refers to the fact that most formulaic sequences contain a word behaving in an unusual way, whether showing grammatical irregularity (e.g. *by and large*) or having an unusual meaning (e.g. *to face the music* meaning “to confront unpleasantness, especially the consequences of one’s errors”). Variability concerns the flexibility of formulaic sequences where some formulaic sequences are entirely invariable, allowing no variations to any constituent part like *hocus pocus* (meaning “nonsense words or phrases used as a formula by quack conjures” or “foolishness or empty pretence used especially to distinguish deception or chicanery”) while others are more flexible, permitting different degrees of variability like *to know SOMETHING like the back of SOMEONE’S hand*. Collocation refers to the lexical cohesion which embraces a relationship between lexical items that regularly co-occur such as *hard word, hard luck, hard facts, hard evidence, etc.* (Halliday and Hassan 1976: 288).

The second common characteristic of formulaic sequences is their function. Most formulaic sequences are situationally dependent, that is, they are only appropriate in certain social situations (Wray 2002).

The third common characteristic of formulaic sequences lies in their meanings, encompassing both the idiomatic/metaphorical meaning as well as the pragmatic meaning. The stereotypical formulaic sequences are the idioms. Wood (1986: 2) defines an idiom as “a complex expression which is wholly non-compositional in meaning and wholly non-productive in form.” Nattinger and DeCarrico (1992: 33) define idioms as “complex bits of frozen syntax, whose meanings cannot be derived from the meaning of their constituents, that is, whose meanings are more than simply the sum of their individual parts”. The definition offered by Irujo (1986a: 288) is “a conventionalised expression whose meaning cannot be determined from the meaning of its parts”. All these definitions emphasise the semantic opaqueness (the idiomatic/metaphorical meaning) of an idiom. However, Cowie (1988a) points out that semantic transparency (and opaqueness) should not be viewed as the essential defining feature of an idiom. While some idioms are totally

non-transparent like *pig in a poke* (meaning “something that is offered in a manner that conceals its true nature or value”), others can be inferred by a little common sense like *the autumn of one’s life* (Wray 2002). Wray (2002: 57) puts emphasis on the fluidity of formulaic sequences, which “allows for them to be compositional on one hand and entirely holistic on another”. Apart from the idiomatic/metaphorical meaning of a formulaic sequence, pragmatic meaning is another important consideration in locating a formulaic sequence. Strings with a literal meaning can be formulaic or non-formulaic, depending on whether they are associated with pragmatic functions (Cowie 1988a). For instance, the string *if I were you/the king...* is formulaic, while a similar string *if I were the one that she really wanted to talk to...* is non-formulaic (Nattinger and DeCarrico 1992: 13).

The last common characteristic of formulaic sequences is the provenance. Formulaic sequences should encompass those that are formulaic at the very start and those that become formulaic (Wray 2002: 59-61). For example, *open sesame* is obviously formulaic at the start that everyone just learns it whole, while *Rice Krispies* shifts from the name of a Brand to referring to all crisp rice over time (Wray 2002: 60).

From these common features, Wray (2002: 62) suggests that formulaic sequences can be best described as lying somewhere on a continuum between the semantically most transparent ones and the semantically most opaque ones as well as between the structurally most fixed ones and the relatively free combinations. This idea of formulaic sequences lying on a continuum between the semantically most transparent ones and the semantically most opaque ones as well as between the structurally most fixed ones and the relatively free combinations is shared by many other researchers like Pawley and Syder (1983), Howarth (1998b), Givón (1989), Sinclair (1987), etc.

Based on rather different criteria, various definitions have been put forward for formulaic sequences by researchers. Nattinger and DeCarrico (1992: 1) define formulaic sequences (“lexical phrases” in their term) as “...conventionalised form/function composites that occur more frequently and have more idiomatically determined meaning than language that is put together each time”. Myles et al. (1999) see formulaic sequences in second language learners as fixed multi-morphemic phrases or sentences which are fluently but perhaps inaccurately produced at times with overextended semantic or pragmatic functions compared with target language norms.



Although many different definitions have been proposed by different researchers, the definition given by Wray is now being widely adopted as a covering definition for the study of formulaic sequence.

*“...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 2002: 9)

It should be noted that Wray's definition also has its problems. Read and Nation (2004) point out that the identification of formulaic sequences based on Wray's (2002: 9) “stored and retrieved whole from memory at the time of use” is quite elusive because the means to store and retrieve the same sequence by one individual may differ from those used by another. Furthermore, the same individual may store and retrieve the same formulaic sequence in one way at one time and another at another time, “depending on a wide range of factors such as changes in proficiency, changes in processing demands, and changes in communicative purposes” (Read and Nation 2004: 25).

Albeit the fact that Wray's (2002) definition has been criticised (Read and Nation 2004), it is still the most widely used throughout the study of formulaic sequences. Thus, this present study will take Wray's definition as the covering term.

Formulaic language can be said to be “ubiquitous” or “pervasive” in daily communications (Yorio 1989; Nattinger and DeCarrico 1992; Wray 2002; Schmitt 2004). However, already in 1998, Myles, Hooper and Mitchell (1998: 324) stated that the study of formulaic language “has not figured prominently in second language acquisition (SLA) research in the last 20 years”, during which period the focus was mainly on the study of creative, rule-governed processes in second language acquisition and on the systematic nature of learners' inter-language.

Formulaic language has long been considered “peripheral to the main body of language” research (Nattinger and DeCarrico 1992: xv). It was either marginalised or completely ignored by researchers in applied linguistics. However, in recent years, it is rapidly becoming a key focus point for research in applied linguistics, and has provided many valuable implications for language teaching (Boers et al. 2006).

Formulaic sequences have their special features and are used dynamically to respond to processing and interactional needs. Some appear in a speaker's speech