



*A Usage-Based Approach to Second Language  
Acquisition of Collocations*

# 基于使用理论视角下的 二语词汇搭配习得研究



高 维 / 著



科学出版社

教育部人文社科青年基金项目（项目批准号：14YJC740024）资助

# 基于使用理论视角下的二语 词汇搭配习得研究

A Usage-Based Approach to Second Language  
Acquisition of Collocations

高 维 著

科 学 出 版 社

北 京

## 内 容 简 介

本书从基于使用理论(The Usage-Based Model)的角度分析二语词汇搭配习得的机制。语言习得的过程是一个动态复杂的系统,学习者所接受的语言输入与认知能力不断交互作用,共同促进语言水平发展。因此,语言输入的频率和结构等特征会在很大程度上影响语言习得的过程和结果。在此理论框架下,本书提出两个可能影响二语词汇搭配习得的重要因素:搭配词的共现频率和搭配的程度。作者假设,搭配词的高频共现会强化该搭配作为整体的心理表征。此外,搭配强度越高,搭配词之间的吸引力越强,该搭配更倾向被视作一个整体储存。为证实该设想,作者开展了两项实验,并且最终数据结果基本印证了作者的研究假设。

### 图书在版编目(CIP)数据

基于使用理论视角下的二语词汇搭配习得研究 = A Usage-Based Approach to Second Language Acquisition of Collocations: 英文 / 高维著. —北京: 科学出版社, 2016.1

ISBN 978-7-03-046301-2

I. ①基… II. ①高… III. ①第二语言-研究-英文 IV. ①H042  
中国版本图书馆 CIP 数据核字(2015)第 267649 号

责任编辑: 阎 莉 王瑞媛 / 责任校对: 蒋 萍  
责任印制: 张 伟 / 封面设计: 铭轩堂

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京京华虎彩印刷有限公司 印刷

科学出版社发行 各地新华书店经销

\*

2016 年 1 月第 一 版 开本: 720 × 1000 B5

2016 年 1 月第一次印刷 印张: 13 1/2

字数: 300 000

定价: 78.00 元

(如有印装质量问题, 我社负责调换)

## 前 言

在语言的使用中,词汇常常以某些固定的模式捆绑在一起出现。这种词与词之间的习惯性结伴使用被称为“词汇搭配”。词汇搭配具有一定程度的因循性和任意性,不能完全按照语法规则来生成。譬如,在汉语中可使用“穿袜子”和“戴手套”,但不会使用“穿手套”或“戴袜子”;在英语中,对“强风”一词的表达,本族语者习惯使用“a strong breeze”或“a strong wind”,也倾向于接受“a stiff breeze”,却很少会说“a stiff wind”。词汇搭配是交际过程中传承下来的习惯用法,是该语言的固有特征之一,并且经常无法跨语言使用。因此,能否正确运用地道的搭配,不仅体现了学习者的语言水平,更体现了学习者对目标语文化的掌握和语用能力。

虽然英国语言学家 J. R. Firth 早在 20 世纪 50 年代就已正式提出“词汇搭配”这一学术概念,但直至 90 年代,国内外语言习得领域对词汇搭配的研究还较少。在计算机语料库问世前,人们对语言结构的研究通常基于直觉,缺乏自然数据支持,研究结果也较为主观和局限。自 20 世纪末起,随着计算机科学的迅猛发展,各类语料库纷纷建立。之后,不断有学者利用语料库对自然语言进行统计和分析,并发现程式性话语并非是前人所认为的少数和边缘的现象,而是无处不在的语言结构,而词汇搭配作为程式性话语的重要构成部分,在语言使用中亦十分普遍。

成年本族语者能够灵活自如地将词汇以地道的方式组合运用;然而,对非本族语者而言,如何恰当地使用目标语的词汇搭配已成为语言习得领域公认的难点之一。迄今为止,大量实证研究已产生丰富的数据,证实即使较高水平的二语学习者在词汇搭配的使用中也存在巨大障碍。不过,造成二语词汇搭配习得困难的原因是什么?如何解决这个二语习得中的难题?我们还需要继续摸索,找出答案。因此,除了分析学习者在词汇搭配使用中的问题之外,还应揭示制约搭配习得的因素,并对学习者习得和使用搭配时的心理表征进行研究。

近年来,国外已有学者(如 Ellis 等, 2009)开始探索母语者对词汇搭配的心理表征,试图通过揭示母语者对搭配的储存和提取方式,进一步认识语言结构的心理表征系统。这一举动不仅能促进词汇搭配研究的深入,揭示搭配习得的机制,还能为目前语言习得理论的转向提供实证数据。不过在二语习得领域,就二语词汇搭配的储存和处理方式,国际上仍有争议,相关实证研究也比较少。

在这样的背景下，本书尝试从基于使用理论（Usage-based Model）的角度分析二语词汇搭配习得的机制。语言习得的过程是一个动态复杂系统，学习者所接受的语言输入与认知能力不断交互作用，共同促进语言水平发展。因此，语言输入的频率和结构等特征会在很大程度上影响语言习得的过程和结果。在此理论框架下，本书提出两个可能影响二语词汇搭配习得的重要因素：搭配词的共现频率和搭配的强度。作者假设，搭配词的高频共现会强化该搭配作为整体的心理表征。此外，搭配强度越高，搭配词之间的吸引力越强，该搭配更倾向被视作一个整体储存。为证实该设想，作者开展了两项实验。实验一采用在线搭配可接受性的任务，主要目的是揭示学习者处理二语搭配时的认知过程。实验二采用书面测试的形式，考察受试对不同搭配的可接受程度的判断。最终数据结果基本印证了作者的研究假设。

本书在理论框架的构建、研究设计和数据分析等各个环节都遇到了许多困难。在此，我首先要向恩师王初明教授致谢，如果没有他不断的启发、鼓励和帮助，我难以坚持并最终完成这项研究；其次，我要感谢好友及同窗王敏，在研究进展不顺利时，她总能在百忙之中及时伸出援手，毫无保留地为我提出宝贵的建议；最后，没有家人的支持，我不可能顺利完成博士学业，也不可能安心从事二语习得研究，在此送上我最真挚的谢意。

由于本人水平有限，本书难免有疏漏不妥之处，我诚恳地希望各位读者能批评指正。

高 维

2015年9月

## List of Abbreviations

ALTE	Association of Language Testers of Europe
AJT	acceptability judgment test
BNC	British National Corpus
CAT	computer-adaptive testing
CBT	computer-based testing
CLAWS	the Constituent Likelihood Automatic Word-tagging System
EFL	English as foreign language
ESOL	English for speakers of other languages
Fe	expected frequency
Fo	observed frequency
FS(s)	formulaic sequence(s)
L1	the first language
L2	the second language
LTUT	Learn Together, Use Together
MI	mutual information
Obj	object
OBL	oblique
P&P	paper and pen
POS	part-of-speech/part of speech
Pe	expected probability
Po	observed probability
QPT	Quick Placement Test
RT	response time
SD	standard deviation
SLA	second language acquisition
Subj	subject
UG	Universal Grammar
V	predicate

## Preface

This book aims to investigate Chinese learners' acquisition of English collocations, focusing on two most commonly addressed issues: a) what is the nature of learners' representation of L2 collocations? b) what factors may shape L2 learners' mental representation of collocations? Within the framework of the usage-based model, we understand collocations as constructions wherein two (or more) words show a strong tendency to co-occur, and thus are represented by learners as lexical chunks. But the degree of chunking varies according to two crucial factors, i.e. frequency and collocation strength. It is assumed that the two factors jointly contribute to L2 learners' collocation representation. In order to achieve the research goals, two empirical studies were conducted. Study 1 drew on a psycholinguistic technique E-prime to reveal how L2 collocations were stored and retrieved in learners' mental lexicon. Utilizing an on-line acceptability judgment test, Study 1 attempted to probe into the accuracy and fluency of learners' recognition of target collocations. Study 2 employed an off-line graded acceptability judgment task, which required subjects to rate the degrees of acceptability of L2 collocations. This study intended to elicit learners' off-line acceptance rates of collocations.

Findings in the empirical studies have largely supported the research hypotheses. In Study 1, three major findings have emerged. First of all, high frequency collocations were recognized more accurately and more quickly than low frequency collocations and non-collocations. Secondly, high collocation strength played a facilitative role in promoting response accuracy rates and times. Thirdly, collocation frequency and collocation strength exhibited interaction effects in modulating the recognition of collocations. As regards the interaction between collocation frequency and collocation strength, it was observed that when collocations were at high-frequency level, collocation strength had a greater magnitude of effects. That is to say, when a collocation is more frequent in the

linguistic exposure, learners are more able to tell the internal cohesiveness of it. When a collocation is infrequent, however, learners are less sensitive to the association strength between the collocation elements. In Study 2, significant main effects of collocation frequency and collocation strength were also detected. First of all, collocations of high frequency were more readily perceived as natural English collocations. Secondly, collocations of high association strength were judged to be more acceptable than loosely-bound ones.

Overall, the findings of the present research show that collocations are stored in L2 learners' mental lexicon in a formulaic tendency. The formulaic nature of a collocation is measured by subjects' processing accuracy, fluency and also the acceptance rates of it. Collocation frequency and collocation strength play significant roles in shaping the mental representation of collocations. The more frequent and stronger a collocation is, the more likely it will be retrieved in a holistic way, and also, the more likely it will be perceived as genuine English collocations. Results from the present study could further lend credence to the usage-based approach to SLA.



# Contents

前言 .....	i
Preface .....	iii
<b>Chapter 1 Introduction .....</b>	<b>1</b>
1.1 What is “Collocation”? .....	1
1.2 The Importance of Studying Collocations in Second Language Acquisition .....	3
1.3 The Purpose of This Book .....	8
1.4 Key Research Questions .....	11
1.5 Organization of the Book .....	11
<b>Chapter 2 A General Overview of Formulaic Sequences .....</b>	<b>13</b>
2.1 Terms and Definitions of Formulaic Sequences .....	15
2.2 Glimpses on the History of Formulaic Sequence Studies .....	18
2.3 Functions of Formulaic Sequences .....	22
2.4 Summary .....	30
<b>Chapter 3 A Thorough Look at Collocations .....</b>	<b>32</b>
3.1 Definitions and Identification of Collocations .....	33
3.2 A Working Definition of Collocation and Related Concepts .....	50
3.3 Previous Studies on Collocations in SLA .....	56
3.4 Summary .....	64
<b>Chapter 4 Towards a Usage-Based Understanding of Collocations .....</b>	<b>66</b>
4.1 A Brief Introduction to the Usage-Based Model .....	66
4.2 Formulating a Usage-Based Framework of Collocation Acquisition and Use .....	81
4.3 L2 Collocation Acquisition: Controversies and Remaining Questions .....	99
4.4 Summary .....	102
<b>Chapter 5 The Empirical Studies .....</b>	<b>104</b>
5.1 General Research Questions and Hypotheses .....	105
5.2 Instruments .....	107
5.3 Study 1 .....	109

5.4 Study 2 .....	142
5.5 Summary of Study 1 and Study 2 .....	150
<b>Chapter 6 Major Findings, General Discussion and Conclusion .....</b>	<b>153</b>
6.1 Major Findings .....	153
6.2 General Discussion .....	160
6.3 Conclusion .....	176
<b>References .....</b>	<b>178</b>
<b>Appendixs .....</b>	<b>194</b>
Appendix 1 Items for the On-Line Experiment .....	194
Appendix 2 Off-Line Acceptability Judgment Test (AJT) .....	199

# List of Tables

2.1	Terms Used in the Literature to Describe Formulaic Sequences and Formulaicity (Wray & Perkins, 2000: 3) .....	15
3.1	Subcategories of Word-like Combinations (Cowie, 1998b: 7) .....	42
4.1	Examples of Constructions (adapted from Goldberg, 2006) .....	70
4.2	The Syntax-Lexicon Continuum (Croft, 2001: 17) .....	86
5.1	Bio-data of Subjects in Study 1 .....	111
5.2	Look-up for Paper and Pen Score (Geranpayeh, 2003) .....	112
5.3	Grouping of Participants Based on Scores in QPT .....	113
5.4	A Comparison of Collocation Lengths Across 4 Conditions .....	122
5.5	Descriptive Statistics of Response Accuracy Rates on Non-collocations and Collocations by Two Proficiency Groups .....	127
5.6	Effects of English Proficiency Level and Word Combination Type on Subjects' Response Accuracy .....	127
5.7	Descriptive Statistics of RTs on Non-collocations and Collocations by Two Proficiency Groups .....	128
5.8	Effects of English Proficiency Level and Word Combination Type on Subjects' RTs .....	129
5.9	Descriptive Statistics of Response Accuracy Rates .....	131
5.10	Results of the Mixed Repeated Measures ANOVA on Response Accuracy Rates .....	132
5.11	Simple Effect Analysis of Collocation Strength by Frequency Level on Response Accuracy Rates .....	135
5.12	Descriptive Statistics of RTs .....	136
5.13	Results of the Mixed Repeated Measures ANOVA on RTs .....	138
5.14	Simple Effect Analysis of Collocation Strength by Frequency Level on RTs .....	141
5.15	Comparisons of Frequency Effect by Proficiency Level .....	141

5.16	Example of the Graded AJT Questionnaire .....	145
5.17	Means and SDs of Acceptance Rates on Non-collocations and Collocations Across the four Different Conditions .....	146
5.18	Results of the Mixed Repeated Measures ANOVA on the Acceptance Rates .....	149

## List of Figures

3.1	Classification of Word Combinations (Cowie, 1998b, 2001).....	43
4.1	Linguistic Constructions in the Specificity/Complexity Continuum (adapted from Roehr, 2008) .....	70
5.1	A Comparison of Response Accuracy on Non-collocations and Collocations .....	126
5.2	Comparison of Response Time on Non-collocations and Collocations.....	128
5.3	Comparison of Response Accuracy Rates of Collocations Across the 4 Conditions.....	130
5.4	Interaction Between Frequency and Collocation Strength .....	134
5.5	A Comparison of RTs of Collocations Across the 4 Conditions.....	137
5.6	Interaction Between Collocation Frequency and Collocation Strength .....	139
5.7	Interaction Between Collocation Frequency and Subjects' English Proficiency .....	140
5.8	Comparing Subjects' Ratings Across Non-collocations and Four Different Types of Collocations .....	148

# Introduction

This book brings Chinese EFL learners' acquisition of English collocations into focus. The primary objectives of the research are two-fold: a) to probe into the mental representations of English collocations by Chinese EFL learners, and b) to explore factors determining the learners' recognition and acceptance of collocations. In this book, it is assumed that collocations are semi-fixed multi-word units, which may be stored in the mental lexicon and retrieved in a holistic tendency. Following the usage-based model, I propose that two major determinants are *collocation frequency* and *collocation strength*. Revealing the significant effects of the factors can shed light on a usage-based approach to second language acquisition (SLA), and help us overcome some difficulties in teaching and learning L2 collocations.

## 1.1 What is "Collocation"?

"Collocation" refers to the habitual co-occurrences and association at the word level, such as *heavy rain*, *hard evidence*, *soft drink*, etc. It is an "aspect of lexical cohesion which embraces a 'relationship' between lexical items that regularly co-occur" (Carter, 1988: 163). As Firth put it, "*you shall know a word by the company it keeps!*" (Firth, 1957: 11), we can fully understand the meaning of a word by looking at its companies. This statement has been one of the most famous and oft-quoted remarks in the literature of linguistic studies. In natural language use, words are not combined into collocations randomly (Erman, 2010), but show strong tendencies to occur in clusters (Kjellmer, 1994). The particular ways in which they go together are a rich and important source of information about language and

about the world we live in (Evert, 2005).

Here are some examples of English collocations.

**Examples:**

a strong breeze/a stiff breeze	a strong wind/* <sup>①</sup> a stiff wind
a tall building/ a high building	a tall boy/*a high boy
make a decision	*do a decision
take advantage of...	*make advantage of...
dark night/broad daylight	*black night/*bright daylight

Native speakers of English tend to use *a stiff breeze* instead of *\*a stiff wind*, whereas *a strong breeze* and *a strong wind* are both idiomatic. Also, both *a tall building* and *a high building* are natural collocations, but people only say *a tall boy*, not *\*a high boy*. It is natural to say *take advantage of* instead of *make advantage of*. It can be seen that collocational patterns are subtle and not easily explainable, however, adult native speakers always have a very good mastery of them (Manning & Schütze, 1999).

Up till now, “collocation” has been considered as a fuzzy concept. In the past literature, based on researchers’ perspectives and domains of study, an abundance of definitions of “collocation” have emerged. Those definitions, as will be reviewed in Chapter 3 in details, fall into two general approaches. The first is the Frequency-Based Approach, according to which collocations are identified based on calculating the actual co-occurrences of words. The second one can be termed as the Phraseological Approach, in which collocations are defined and identified purely according to the semantic and syntactic properties. The meaning attached to the term “collocation” depends basically on the particular application being used for (Anagnostou & Weir, 2006). To attain a clear understanding of the term, in the confines of the present research, the Frequency-Based Approach has been adopted to define and identify collocations. In the present research, “collocation” is understood as “grammatically restricted co-occurrences of at least two words,

① Items with “\*” are unacceptable word combinations.

which have a greater-than-chance likelihood to appear in a text”.

## 1.2 The Importance of Studying Collocations in Second Language Acquisition

Collocations, as a subtype of formulaic sequences (FSs), have long been going unnoticed for decades. It is generally acknowledged that formulaic sequences constitute the core of idiomaticity. Native speakers of a language are capable of telling very subtle differences among a multitude ways of wording ideas, and they could tell which expression is the most natural one and which is the odd one, regardless of grammaticality. However, the most advanced L2 learners may fail to achieve this “native-like accuracy” and “native-like fluency” even after years of studying a language (Pawley & Syder, 1983). It is believed that a good knowledge of formulaic sequences is the key to solve this problem, because learning a language is not just a matter of knowing the rules, but also knowing a habit or a culture (e.g. Pawley & Syder, 1983; Kecskés, 2002). Formulaic sequences are the well-established ways of habitually expressing an idea, which embodies social acknowledgement and background knowledge. They do not only consist of fixed expressions, such as idioms and proverbs, etc., but also semi-fixed multi-word units, such as collocations.

However, among all categories of conventional lexical combinations, idioms have attracted extensive attention in linguistic theory and description (Sinclair, 2004b). They have been taken as the key to the door of a speech community and native-like language proficiency; whereas other subtypes of formulaic sequences, especially collocations, have been treated with considerable neglect (Bartsch, 2004). Thus far, systematic research of collocations is new and far from exhaustive (Bartsch, 2004). Considering that collocations are a ubiquitous phenomenon and collocational knowledge is a prerequisite for successful communication (Nakata, 2007), and they present a severe burden for L2 learners, the present research strongly argues that collocations deserve more attention and investigation than what the previous literature has known.



First of all, collocations are everywhere in natural language use. Formulaic sequences including collocations used to be considered as a marginal phenomenon. In recent years, however, the advent of corpus techniques has enabled us to extract patterns of real language use. There are studies alleging that collocations are a ubiquitous phenomenon in language structures, and collocations make up a large proportion of multi-word expressions. For instance, the query of UK-part of the search engine Yahoo showed that collocations are much more frequent than idioms (Handl, 2008). Cowie (1992; cf. Pawley, 2007) searched two articles in *The Times* and found that about 35 to 45 percent of all the sequences of a given structural type consisted of restricted collocations. In a similar vein, in an analysis of over 5,000 verb-noun combinations in a written 240,000-word corpus, over a third of the combinations were found to be collocations (Howarth, 1996). Kjellmer (1987: 140) also claimed that “in all kinds of texts, collocations are indispensable elements with which our utterances are very largely made”. This suggests that we can no longer treat collocations as a minor part in linguistic studies.

Secondly, apart from the pervasiveness of collocations, what makes studies on this issue necessary is that collocations are a major stumbling block in the L2 learning process. Competence of collocations belongs to native speaker’s intuition; and for adult native speakers, communication is hardly impeded by collocational knowledge. In stark contrast, for non-native speakers, learning collocations remains tricky and unmanageable for quite a long time (Higuchi, 1999). It has been substantiated that non-native speakers’ collocational knowledge largely lags behind their general vocabulary knowledge and mastery of grammar rules, and even very advanced learners tend to usually produce inappropriate or unacceptable collocations (McCarthy, 1990: 13). Howarth (1996), for example, conducted a book-length study to examine the use of conventional lexical collocations in written academic performance in English by native and non-native university students. Focusing on “*verb+object*” collocations, he found that restricted collocations presented more problems for non-native writers than idioms did. He observed that the collocational errors made by non-native writers led to “a lack of