



基于语料库的中国英语学习者 词语搭配行为研究

王海华 著

Collocational Behavior of Chinese EFL Learners

Investigating the Collocational Behavior of Chinese EFL Learners: A Corpus-Based Approach

A Corpus-Based Approach

外语教学与研究出版社

FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

基于语料库的中国英语学习者 词语搭配行为研究

**Investigating the Collocational Behavior of Chinese EFL Learners:
A Corpus-Based Approach**

王海华 著



外语教学与研究出版社

FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

北京 BEIJING

图书在版编目(CIP)数据

基于语料库的中国英语学习者词语搭配行为研究 / 王海华著. — 北京: 外语教学与研究出版社, 2009. 3

(外研社高等英语教育学术文库)

ISBN 978-7-5600-8229-5

I. 基… II. 王… III. 英语—口语—研究 IV. H319.9

中国版本图书馆CIP数据核字 (2009) 第033474号

出 版 人: 于春迟

责任编辑: 陈 静

封面设计: 覃一彪

出版发行: 外语教学与研究出版社

社 址: 北京市西三环北路19号 (100089)

网 址: <http://www.fltrp.com>

印 刷: 北京国工印刷厂

开 本: 650×980 1/16

印 张: 16.75

版 次: 2009年10月第1版 2009年10月第1次印刷

书 号: ISBN 978-7-5600-8229-5

定 价: 39.90元

* * *

如有印刷、装订质量问题出版社负责调换

制售盗版必究 举报查实奖励

版权保护办公室举报电话: (010)88817519

物料号: 182290001

前言

外语学习者虽然能够说出语法上正确的句子，但是他们所说的英语有时听起来还不是“地道”的英语，原因之一是由于学习者在交流的过程中不能准确地使用大量的搭配结构。词语搭配现象是词汇行为的基本模式，是词汇选择的基本机制。作为词汇知识的一个重要组成部分，词语搭配行为越来越受到研究者的重视，对其研究也越来越深入。研究表明，外语学习者在使用词语搭配方面存在着较大的问题，而语料库数据（本族语语料库和学习者语料库）为学习者词语搭配行为的研究提供了有力的数据支持，基于语料库的词语搭配行为的研究使人们重新审视语言中的搭配模式及其重要性。

本研究以中国学习者英语语料库（CLEC）和英语国家语料库的部分语料（SBNC）为主，同时使用了BROWN和LOB两个语料库，在作对比分析时使用的是PFR People's Daily汉语语料库，检索工具使用的是WordSmith Tools和ConcApp。基于上述语料库和检索工具，本研究从三个方面来探讨中国学习者的词语搭配行为。

首先，本研究从学习者在词语搭配方面的错误着手，研究中国学习者的搭配行为特点。中国学习者英语语料库对六类搭配错误进行了标注：CC1为名词与名词的搭配错误，CC2为名词与动词的搭配错误，CC3为动词与名词的搭配错误，CC4为形容词与名词的搭配错误，CC5为动词与副词的搭配错误，CC6为副词与形容词的搭配错误。从搭配错误的整体分布来看，CC3所占比例最高，为54.9%；CC6所占比例最少，为1.8%。其余的错误所占比例分别是：CC1为11.8%，CC2为12.2%，CC4为16.1%，CC5为3.2%。本研究对其中的四种搭配错误，即CC1、

CC2、CC3 和 CC4 进行了详细的分析。为了使课程标准的制订者、教材的编写者以及教师，能对中国英语学习者使用语言的机制、搭配错误的特点有详细地了解，从而更有针对性地制订课程标准，编写教材并进行教学，本研究将学生的词语搭配错误进行详细归类分析。结果显示，中国学生搭配错误的主要特点是字面翻译、近义词误用、替代、过于简化、冗余、误选和迂回表达。其中，以近义词误用和字面翻译最为明显，汉语对搭配错误的形成影响最大。

第二，以词语搭配的一种特殊现象——语义韵为切入点来研究学习者的搭配行为。此项研究以具有明显消极语义韵的 CAUSE 和具有明显积极语义韵的 PROVIDE 为重点，将 CLEC 和 SBNC 两个语料库中所出现的 CAUSE 和 PROVIDE 搭配词的语义倾向进行对比分析。在 CAUSE 的 10 个高频搭配词中，有四个词项一致，它们是 change(s), problem(s), harm 和 great(er, est)。其中 problem(s) 和 harm 是具有明显消极涵义的词；change(s) 和 great(er, est) 没有明显的语义倾向，是中性词。在 PROVIDE 的 10 个高频搭配词中，有三个词项一致，它们是 more, information 和 opportunity(ies)。其中 information 和 opportunity(ies) 是具有明显积极涵义的词，more 是个中性词。本研究分别对 CAUSE 与 change(s)、great(er, est)，PROVIDE 与 more, information, opportunity(ies) 共现的语境进行了对比分析。研究表明：在 CAUSE 与 change(s) 和 great(er, est) 共现时，中国英语学习者所使用的语境过多地表现出积极语义韵，少用了其典型的消极语义韵；在 PROVIDE 与 more, information, opportunity(ies) 共现的语境中，中国英语学习者同本族语者所表现的语义倾向一致。该研究还对语义韵的功能进行了探讨。因为有意识地运用语义韵的冲突，能传达作者的真实态度，有时会达到讽刺的效果；无意识造成语义韵的误用，会使正常的交际受阻，有时会导致语用语言失误。

第三，二语习得主要是对第二语言学习者的中介语能力及其发展特点进行研究。前两项研究是把中国英语学习者作为一个整体，来研究其词语搭配使用的特点；而此项研究以衡量学习者语言水平高低的三个群体（高中、大学非英语专业和大学英语专业的学习者）为基础，以扩大词搭配的使用为切入点，来研究中国学生英语搭配知识的发展特点。该

项研究对中国英语学习者使用扩大词的总体情况进行了分析，并重点从搭配的分类、形符及其准确率方面分析六个提升词（badly, deeply, greatly, so, very, very much）搭配的使用。结果显示，相对低水平的学生过多使用“万能词”——so, very, very much 的搭配。这些学生过多使用这些词，主要是用来弥补他们词汇的不足或是避免错误。就其所使用这三个“万能词”的准确率及复杂性而言，与学生的英语水平成正比：即水平越高，所使用搭配的准确率也越高，搭配结构也越复杂。在使用 badly, deeply, greatly 的搭配词项时，主要差异出现在高中和大学英语专业学习者这两个群体中，大学非英语专业和大学英语专业学习者这两个群体中。中国英语学习者在使用这三个提升词的搭配时，其准确率呈现出“U”形的态势，即在高中群体中准确率较高，在大学非英语专业群体中有所下降，在大学英语专业群体中准确率又升高。

最后，在对本研究作总结的同时，作者也就本研究的局限性及今后研究的发展取向进行了探讨，并在以上研究的基础上，讨论了该研究在教学中的意义，即如何根据中国英语学习者的具体特点来进行词语搭配的教学。

本研究是对中国学习者英语词语搭配行为的一个尝试性探索，无论是研究的角度，还是研究的方法，都有别于国内外的相关研究。

王海华

2009年4月

Acknowledgements

I feel deeply indebted to many people who have become involved in my interest and helped me in different ways.

First and foremost, I am indebted to my supervisor, Professor Wang Tongshun. During the three years of Ph.D. study in Shanghai Jiao Tong University, he has always been a patient, kind and skilful supervisor and mentor, as well as a constant source of support and encouragement. Professor Wang has provided insightful comments as well as suggestions.

I would like to extend my heartfelt thanks to Professor Liu Runqing and Professor Chen Guohua, who have been very generous with their time and ideas, and whose ideas on how to do research have greatly influenced me.

I am deeply grateful to Professor Yang Huizhong and Professor Wei Naixing, who ushered me into the discipline of corpus linguistics.

During my one-year study at the University of Nottingham, UK, I met a few helpful experts and friends who deserve my special thanks. They are, to name but a few, Professor Carter who generously lent me the book *Reading Concordance*, which was otherwise unavailable; Dr. Schmitt who sent me copies of his papers inaccessible at that time; a few English friends who helped me a lot in getting native speakers' intuitions; and Professor Hunston who gave me some useful suggestions when I presented my paper at BAAL annual conference in London.

Deeply felt thanks are also due to John Redenbach, a foreign teacher in Dalian Maritime University, who has proofread my book.

Appreciation is also expressed to my husband, Liu Zhipeng, without whom none of this would have been possible. He has always been an

unfailing source of support and encouragement, especially in times of difficulty and frustration.

Any remaining shortcomings you may perceive are, of course, my own responsibility.

List of Tables

Table 4.1	Raw frequency of lexical collocational errors in HS, NM and EM
Table 4.2	Normalized frequency of lexical collocational errors in HS, NM and EM
Table 4.3	Idiosyncratic redundancy of “noun + noun”
Table 4.4	Idiosyncratic assumed synonyms of “noun + noun”
Table 4.5	Idiosyncratic over-simplification of “noun + noun”
Table 4.6	Collocates of <i>knowledge</i>
Table 4.7	Collocates of <i>society</i>
Table 4.8	Collocates of <i>success</i>
Table 4.9	Collocates of <i>ability(ies)</i>
Table 4.10	Normalized frequency of seven adjectives in CLEC, SBNC, BROWN and LOB
Table 5.1	Raw frequency of CAUSE in CLEC and SBNC
Table 5.2	Normalized frequency of CAUSE in CLEC and SBNC
Table 5.3	Collocates of CAUSE with top 58 I value in SBNC and CLEC
Table 5.4	Top 10 frequent collocates in CLEC and SBNC
Table 5.5	Distribution of connotational meanings when <i>change(s)</i> co-occurs with CAUSE
Table 5.6	Distribution of connotational meanings when <i>great(er, est)</i> co-occurs with CAUSE
Table 5.7	Frequency of PROVIDE in CLEC and SBNC
Table 5.8	Collocates of PROVIDE in CLEC
Table 5.9	Top 10 frequent collocates of PROVIDE in CLEC and SBNC
Table 6.1	Overall frequency of individual maximizers in HS, NM and EM

Table 6.2	Overall frequency of individual boosters in HS, NM and EM
Table 6.3	Normalized frequency of amplifier tokens in HS, NM and EM
Table 6.4	Normalized frequency of booster tokens in HS, NM and EM
Table 6.5	Tokens of six boosters in HS, NM and EM
Table 6.6	Summary of <i>badly</i> collocates
Table 6.7	Summary of <i>deeply</i> collocates
Table 6.8	Summary of <i>greatly</i> collocates
Table 6.9	Summary of <i>very</i> collocates
Table 6.10	Combination patterns of <i>so</i> in HS
Table 6.11	Combination patterns of <i>so</i> in NM
Table 6.12	Combination patterns of <i>so</i> in EM
Table 6.13	Typical patterns of <i>so</i> in HS, NM and EM
Table 6.14	Distribution of <i>very much</i> in HS, NM and EM
Table 6.15	Accurate use of <i>very much</i> in HS, NM and EM
Table 6.16	Recurrent combinations of <i>very much</i> in HS
Table 6.17	Recurrent combinations of <i>very much</i> in NM
Table 6.18	Recurrent combinations of <i>very much</i> in EM

List of Figures

- Figure 2.1 Interlanguage
- Figure 2.2 Error patterns predicted by the Ontogeny Model
- Figure 2.3 Contrastive Interlanguage Analysis
- Figure 3.1 Possible applications of a corpus
- Figure 3.2 Specific criteria for designing computer learner corpora
- Figure 3.3 Span
- Figure 4.1 Distribution of the six types of collocational errors in CLEC
- Figure 4.2 Distribution of the six types of collocational errors in HS
- Figure 4.3 Distribution of the six types of collocational errors in NM
- Figure 4.4 Distribution of the six types of collocational errors in EM
- Figure 4.5 Typical collocates of the five adjectives from SBNC
- Figure 4.6 Collocates of *job* in Chinese
- Figure 4.7 Collocates of *job* as retrieved from CLEC
- Figure 4.8 Collocates of *job* as retrieved from the NS corpora
- Figure 4.9 Summarization of causes and categories of collocational errors
- Figure 6.1 Subtypes of intensifiers
- Figure 6.2 Distribution of different patterns of *so* in HS, NM and EM
- Figure 6.3 Distribution of amplifiers and accuracy rate of the five boosters
as tagged on y-axis
- Figure 6.4 Schema of U-shaped behavior
- Figure 7.1 Collocational grid of *big*, *large*, *great* and *huge*

List of Abbreviations

BNC	the British National Corpus
BROWN	the Brown corpus
BWF	Bundeswettbewerb Fremdsprachen Corpus
CA	Contrastive Analysis
CAH	Contrastive Analysis Hypothesis
CEA	Computer-Aided Error Analysis
CET	college English test
CIA	Contrastive Interlanguage Analysis
CLC	computer learner corpora
CLEC	the Chinese Learner English Corpus
COBUILD	the Collins Birmingham University international language database
DDL	data-driven learning
EA	error analysis
EFL	English as a foreign language
EM	English majors
GCE	General Certificate of Education
HS	high school students
III	illustration, interaction, induction
IL	Interlanguage
L1	first language
L2	second language
LOB	the Lancaster-Oslo/Bergen corpus
LOC	Louvain Corpus of Native English Essays
NM	non-English majors
NL	native language

NNS	non-native speaker
NS	native speaker
SBNC	sub-corpus of the British National Corpus
SLA	second language acquisition
TL	target language
UNI	university corpus

Table of Contents

List of Tables.....xiii

List of Figures xv

List of Abbreviationsxvii

Chapter 1 Introduction..... 1

 1.1 Background 1

 1.2 Research Objectives 3

 1.3 Outline of the Book..... 4

**Chapter 2 Contrastive Analysis, Language Transfer,
 Error Analysis and Interlanguage 6**

 2.1 Introduction 6

 2.2 Contrastive Analysis Hypothesis..... 6

 2.3 Language Transfer 8

 2.4 Error Analysis 11

 2.4.1 EA in General..... 11

 2.4.2 Mistakes and Errors..... 12

 2.4.3 Stages Involved in EA..... 13

 2.4.3.1 Recognition..... 13

 2.4.3.2 Description 14

 2.4.3.3 Explanation 14

 2.4.4 Limitations of EA..... 15

 2.5 Interlanguage 15

 2.6 Contrastive Interlanguage Analysis and Computer-Aided
 Error Analysis..... 19

2.6.1 Contrastive Interlanguage Analysis.....	19
2.6.2 Computer-Aided Error Analysis	20
2.7 Conclusion	21
Chapter 3 Corpus Linguistics and Collocational Studies	22
3.1 Corpus Linguistics: An Overview.....	22
3.1.1 Development of Corpus Linguistics.....	24
3.1.2 Corpora Applied: Some Areas of Corpus-Based Research	26
3.1.3 Pedagogical Implications	29
3.2 Computer Learner Corpora	30
3.2.1 Computer Learner Corpora in General.....	30
3.2.2 Learner Corpora and IL Analysis	33
3.2.3 Learner Corpora and Language Teaching.....	34
3.3 Corpora Employed in the Present Research.....	36
3.3.1 Sub-Corpus of the British National Corpus.....	36
3.3.2 Chinese Learner English Corpus.....	37
3.4 Limitations of Corpora.....	38
3.5 Studies into Collocations.....	39
3.5.1 Word Knowledge Framework.....	39
3.5.2 Collocational Knowledge.....	42
3.5.2.1 Defining Collocation	42
3.5.2.2 Measuring Collocation	45
3.5.3 Importance of Collocational Knowledge.....	48
3.6 Collocation and Language Teaching.....	51
3.7 Conclusion	55
Chapter 4 Collocational Problems of the Chinese EFL Learners	56
4.1 Introduction.....	56
4.2 Lexical Errors.....	57
4.3 Recent Corpus-Based Studies into Collocational Problems of the Learners of English	62
4.4 Method.....	66
4.5 Results and Analysis.....	66

4.5.1 Results	66
4.5.2 Analysis of the Collocational Pattern of “Noun + Noun”	69
4.5.2.1 Calque.....	71
4.5.2.2 Redundancy	72
4.5.2.3 Mis-Selection	73
4.5.2.4 Assumed Synonym	74
4.5.2.5 Over-Simplification	75
4.5.2.6 Circumlocution	75
4.5.3 Analysis of the Collocational Pattern of “Noun + Verb”	76
4.5.3.1 Calque.....	77
4.5.3.2 Assumed Synonym	79
4.5.3.3 Substitution.....	80
4.5.4 Analysis of the Collocational Pattern of “Verb + Noun”	80
4.5.5 Analysis of the Collocational Pattern of “Adjective + Noun”	87
4.5.5.1 Assumed Synonym	88
4.5.5.2 Calque.....	90
4.5.5.3 Mis-Selection	90
4.6 Discussion	91
4.7 Conclusion	98

Chapter 5 Semantic Prosodies of CAUSE and PROVIDE:

A Contrastive Study	99
5.1 Introduction	99
5.2 Semantic Prosody and Related Studies.....	100
5.3 Method.....	106
5.4 Study of CAUSE.....	106
5.4.1 Dictionary Definitions	106
5.4.2 Overall Frequency of CAUSE in CLEC and SBNC.....	108
5.4.3 Data in CLEC and SBNC.....	109
5.4.4 CAUSE Collocating with <i>Change(s)</i>	113
5.4.4.1 Data from CLEC	113
5.4.4.2 Data from SBNC	115
5.4.5 CAUSE Collocating with <i>Great (er, est)</i>	118

5.4.5.1 Data from CLEC	118
5.4.5.2 Data from SBNC	119
5.4.6 Comparisons between CLEC and SBNC.....	122
5.5 Study of PROVIDE.....	126
5.5.1 Dictionary Definitions	126
5.5.2 Overall Frequency of PROVIDE in CLEC and SBNC.....	127
5.5.3 Data in CLEC and SBNC.....	128
5.5.4 Comparisons between CLEC and SBNC.....	131
5.6 Discussion	135
5.6.1 Reasons for Differences and Similarities	135
5.6.2 Functional Effects of Semantic Prosodies	138
5.7 Conclusion	144
Chapter 6 Development of Amplifier Collocational Knowledge: A Cross-Sectional Study	145
6.1 Introduction	145
6.2 Amplifier.....	146
6.3 Research Related to Amplifier Collocations	148
6.4 Method.....	152
6.5 Results	153
6.5.1 Overall Frequency of Amplifiers in the Corpora	153
6.5.2 Booster Collocations	156
6.5.2.1 Analysis of <i>Badly</i>	158
6.5.2.2 Analysis of <i>Deeply</i>	159
6.5.2.3 Analysis of <i>Greatly</i>	160
6.5.2.4 Analysis of <i>Very</i>	161
6.5.2.5 Analysis of <i>So</i>	162
6.5.2.6 Analysis of <i>Very Much</i>	165
6.6 Discussion	167
6.7 Conclusion	175
Chapter 7 Summary and Implications.....	177
7.1 Major Findings.....	177
7.2 Implications.....	180