

L2 Lexical Co-selection Patterns and  
Teaching Design:  
A Corpus-based Study

# 二语词语的共选型式 研究与教学设计

陆军 等著



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

L2 Lexical Co-selection Patterns and  
Teaching Design:  
A Corpus-based Study

# 二语词语的共选型式 研究与教学设计

陆军 等著



上海交通大学出版社  
SHANGHAI JIAO TONG UNIVERSITY PRESS

## 内容提要

本书以语料库驱动的研究思想为指导、以外语教师语料库素养培育为目标、以词语共选型式的教学设计为抓手、以我国现行外语教育人才培养模式为契机,首创性地开发了“学习者语料库驱动的二语教学设计”的教学研究范式,力图为打通语料库研究成果向教学实践应用转化的“最后一公里”提供方法和路径。本书主要涵盖了语料库词语行为研究的核心概念、技术方法,以及语料库驱动的教学设计范式和应用样例。本书适用于教学一线的大中小学外语教师、语言学方向的研究生等,特别可为外语学科方向的教育硕士提供毕业论文写作指导。

## 图书在版编目(CIP)数据

二语词语的共选型式研究与教学设计/陆军等著

—上海:上海交通大学出版社,2022.1

ISBN 978-7-313-24751-3

I. ①二… II. ①陆… III. ①第二语言—词汇—教学设计 IV. ①H003

中国版本图书馆CIP数据核字(2022)第029989号

## 二语词语的共选型式研究与教学设计

ERYU CIYU DE GONGXUAN XINGSHI YANJIU YU JIAOXUE SHEJI

著者:陆军等

出版发行:上海交通大学出版社

邮政编码:200030

印制:江苏凤凰数码印务有限公司

开本:710mm×1000mm 1/16

字数:264千字

版次:2022年1月第1版

书号:ISBN 978-7-313-24751-3

定价:98.00元

地址:上海市番禺路951号

电话:021-64071208

经销:全国新华书店

印张:14.5

印次:2022年1月第1次印刷

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系

联系电话:025-83657309

当下,人类已迈入大数据时代。作为一种新型生产要素,大数据已经悄然登场亮相,对我们的学习、工作和生活产生了巨大影响。语料库建设和语料库软件开发是大数据技术革命在语言研究等领域的具体体现,对语言教学、学习和研究产生了根本性影响。在这样的背景下撰写本书,目的很明确,就是尝试促进语料库研究成果向教学实践转化,试图帮助语言教师和学习者分享到语言大数据的红利。

作为一名长期奋斗在教学一线的外语教师,我切身感受到外语教学和学习的不易和艰辛。我们花费大量时间和精力教授词汇、操练语法规则,但学生总会有很多表达不符合目标语习惯,如 *learn knowledge*\*<sup>①</sup>、*master meaning*\*、*grasp skill*\*、*life expectancy is ... years old*\* 等;我们记忆了大量近义词,但是很难区分其用法上的差异,如 *master skills vs grasp skills*\*、*reject the idea vs refuse the idea*\*、*tolerate high temperature vs endure high temperature*\* 等表达地道与否,也难以区分命题意义接近但态度意义近乎对立的表达,如 *persistent effort* (趋于“厌烦、反对”) vs *constant effort* (趋于“欣赏、赞成”) 等。这些特征对语言表达的准确性和地道性至关重要,是影响二语学习和表达的关键因素之一。更为重要的是,此类语言现象普遍存在,但在二语教学和学习中难以区分。

作为一名长期从事语料库语言学研究的学者,我切身感受到语料库研究在反映语言使用特征、揭示二语学习规律,以及解决二语学习困难等方面具有空前的便利性。试想一下:如何区分汉语中“二”和“两”的用法?这两个词都与英语单词 *two* 高频对应,习惯性地被视为同义词或近义词。平行语料库数据显示, *two cars*、*two bottles of wine* 和 *two rooms* 分别倾向于翻译为“两台车”“两瓶酒”和“两间房”,而基本不译为“二台车\*”“二瓶酒\*”和“二间房\*”。再试想一下,

① 本书中的星号“\*”如无特殊标注,均用于表示非典型表达。

“怀疑”一词表示“相信”还是“不相信”？大多数读者的第一反应可能是后者，倾向于拒绝前者。然而，“怀疑里面有猫腻”之类的表达（此处的“怀疑”趋于表达“相信”之义）在汉语语料库中也高频出现，我们在理解上基本不会产生歧义。上述情形分别涉及普遍存在的同义词现象和一词多义现象，反映了传统语言学中的典型意义模型，即同义词（*thesaurus*）和词典（*dictionary*）（Sinclair 1991: 6-7）。不过，依靠传统的词典和同义词词典不容易说清楚“怀疑”之类的多义词在什么情形中表达何种意思，也难以区分“二”vs“两”之类的近义词的用法特征。作为母语者，我们也很难显性描述这些词语的具体使用规则或搭配规律。这意味着即使依靠本族语者语言直觉也未必能够解决好二语教学和学习中的词语搭配或行为问题。值得庆幸的是，这些问题都能够在语料库语言学中得到很大程度的解决（参见本书第5、6、7等章；Tognini-Bonelli 2001；陆军、卫乃兴 2012；陆军 2019）。

近年来，基于语料库的二语学习语言研究已经取得长足的发展。在国际上，格兰杰（Granger 1996, 1998a）首次建立大规模学习者英语语料库，并提出中间语对比分析法（*contrastive interlanguage analysis*, CIA）。这一里程碑式的成果催生了一大批基于语料库的学习者语言研究，如格兰杰（Granger 1998b）、内塞尔豪夫（Nesselhauf 2003, 2005）、托诺（Tono 2004）、吉尔坎（Gilquin 2000/2001, 2008a, b）、帕科（Paquot 2013, 2019）等，研究数量高达2 000篇（部）（Tono 2018）。在我国，桂诗春、杨惠中（2003）开发和建设了“中国学习者英语语料库”，开创了我国学习者英语语料库研究的先河。大量语料库调查发现，中国英语学习者在词语搭配、类联接、语义趋向、语义韵等词语（共选）型式特征方面都与英语本族语存在偏离，普遍倾向于依靠母语与目标语对应词的联系选择词语和语法结构（参见濮建忠 2003；娄宝翠 2004；李文中 2005；卫乃兴 2006；陆军、卫乃兴 2014；陆军 2018等）。与此同时，何安平（2004, 2009, 2010, 2013）和何安平、许家金、张春青（2020）等一直致力于使用语料库开展外语教学研究。尽管如此，真正意义上基于语料库研究成果的课堂教学实践还非常有限，基于学习者词语共选型式调查结果的实践教学尤为罕见。由此可见，语料库研究成果向教学实践转化还有一段距离。

如何才能解决这一问题呢？新近提出的外语教师语料库素养理念（*corpus literacy*）为我们提供了思想指导。所谓语料库素养，就是要求教师具有能够使用语料库语言学技术来研究语言并促进学生语言发展的能力（Heather & Helt 2012）。卡利斯（Callies 2019）将教师语料库素养概括为四点，即不仅要理解语料库的基本概念、掌握语料库的基本操作、能够诠释语料库数据，还要利用语料库开发教材和教学活动。其中，前三点与希瑟、黑尔特（Heather & Helt 2012）的第

一部分内容相当,即具有语料库研究能力,而第四点则与第二部分相当。由此可见,教师若要能够有效地开展语料库教学研究,不但需要具有利用语料库开展教学的能力(如能够使用好语料库教学案例),还要具有语料库语言研究能力(如能够开展基于语料库的目标语、学习者语言研究),更需要在语料库语言研究的基础上开展语言教学(如根据语料库研究揭示的语言特征和学习规律开发学习任务、设计教学活动等)。这三者道破了上述“一段距离”的问题实质:需要培养具有上述三方面能力的语言教师,三者缺一不可。

基于上述分析,本书的撰写主要作了以下三方面的思考。

**(1) 本书以外语教师语料库素养培育为目标,在内容构成上尽可能满足教师三方面语料库能力培养的需要。**本书第1章至第3章介绍了语料库语言学词语共选型式研究的基本概念。第4章介绍了二语共选型式研究的基本语料库技术,力图满足语料库素养的前两个要求。第5章到第11章都由两大部分构成:前半部分分别从实词和语法词的共现型式阐释本族语和学习者语料库数据,揭示语言使用特征和学习规律,力图满足培养教师开展语料库语言研究的能力的需要;后半部分都是在相应语料库调查的基础上进行教学设计探讨、创设相关教学任务和活动,本书称之为“学习者语料库驱动的教学设计”或“基于语料库的教学设计”。这部分也是本书的主要特色所在,不仅仅充当语料库教学案例的角色,更重要的是帮助教师知其然并知其所以然,即在开展学习者语料库语言研究的基础上进行教学设计和实践探索。一言蔽之,上述内容构成为力图满足语料库素养培养的需求。

**(2) 本书以词语共选型式的教学设计为抓手,诠释基于语料库的二语教学研究思想。**语料库研究的方法有“语料库参照的方法”(corpus-informed/illustrated approach)与“基于语料库的方法”之区分(corpus-based approach)(McEnery & Hardie 2012)。前者具有数据真实性优势,但可能有很大的偶然性,后者除了具有真实性优势以外,还具有比较好的典型性和代表性,有助于较为客观地揭示典型语言型式和使用规律。典型词语共选型式是真实语言使用中的主要意义单位,需要采用基于语料的研究路径才能发现(Sinclair 1991, 1996; Hoey 2005)。不过,需要指出的是,即使是基于语料库的研究成果,也未必可以直接用于教学活动。这是因为二语学习有其自身的规律。就词语型式学习而言,学习者的主要困难存在于双语不对应部分,而很多对应部分可能不需要专门讲授或学习(Bahns 1993)。这意味着并非所有的典型语言结构都是二语教学的重心。相比之下,二语学习或使用中存在主要困难的部分,特别是常规教学方法难以克服的难点,更应该是二语教学的主要关注点。肯普(Kemp 1971)等强调,教学设计活动要开展“系统性需求和问题分析”。相应地,发现二语学习和使用困难是

相关教学设计需求分析的主要目的。基于此,我们提出了基于语料库的二语教学设计思想。其核心内容包括基于语料库开展二语学习困难和需求分析、基于分析结果开展二语教学设计。一定意义上讲,基于语料库的二语教学设计旨在把语料库研究与二语教学实践应用进行系统整合,而不是简单地利用语料库数据开展教学设计,也不同于把语料库调查结果直接拿来开展教学设计。

(3) 本书以我国现行外语教育人才培养模式为契机,促进语料库语言学研究成果向外语教学实践转化、推广。真正意义上践行语料库素养理念和二语教学设计思想、把语料库语言学研究成果转化为教学实践应用是一项要求很高的使命,很难在一线外语教师中直接推广。不过,我国研究生教育专门设置了教育硕士专业学位(Ed. M),主要目的是培养具有现代教育观念、具备较高理论素养与实践能力的骨干教师。其中,外语学科教学方向的教育硕士培养已经成为当下中小学外语师资的主要来源。他们在攻读学位期间需要专门培养教学与研究能力。其中,开展教学设计研究对提高他们的教师素养和课堂教学能力都具有重要价值,是必不可少的环节。在研究生教育阶段引导他们开展基于语料库的二语研究(有助于培养学习者语料库语言学研究能力)、利用研究发现指导教学设计(有助于培养他们在科研成果指导下进行教学实践的能力),不失为培养连接语料库语言学研究与外语教学实践的桥梁人才、解决诸多中小学外语词汇语法教学瓶颈问题的有效路径。本书正是为培养上述桥梁人才、满足当下的外语教学需求准备条件。

本书一共由11章构成。其中,第1章至第5章由陆军完成,第6章和第7章由陆军、卫乃兴合作完成,第8章由陆军、官丽丽合作完成,第9章由陆军、吴慧思合作完成,第10章和第11章由杨卉卉、唐少华、刘建娟、陆军合作完成。第5章至第11章的共同之处在于:首先基于语料库开展学习者语言特征和规律调查研究,然后以语料库研究结果为指导深入开展相关词汇语法型式的教学设计,力图做到研究与应用相结合。这些章节的具体研究过程都由教学一线的老师全程参与,在保证一定理论高度的同时也确保其在实践中的可操作性,可为学习者语料库教学研究方向的教育硕士毕业论文写作提供参考。

拙作付梓之际,心存无限感激。由衷感激北京航空航天大学卫乃兴教授的谆谆教诲。卫乃兴教授的科学数据观和语料库语言学思想一直引领和指导着我探索语言的奥秘,特别是利用语料库开展外语教学研究令我受益匪浅。本书中的研究问题提出、研究方法论证、语料库数据分析以及教学任务设计等都得益于卫乃兴教授直接或间接的指导。需要着重指出的是,本书并非把语料库数据或研究成果直接用于教学,而是根据学习者语料库数据调查首先开展需求分析,然后设计教学任务,这种“学习者语料库驱动的二语教学设计”思想尤其得益于卫

乃兴教授数据驱动理念的启迪。本书的撰写还得益于上海交通大学语言文字工程研究所的各位兄弟姐妹们多年来的合作和帮助；得到了扬州大学外国语学院俞洪亮教授、秦旭教授、王金铨教授、于建华教授、周卫京教授、王洪刚教授和缪海涛教授等领导 and 同事们的帮助和支持。上海交通大学出版社编辑部的老师们对本书的出版给予很大的支持。在此一并表示衷心感谢！

本研究为江苏省社科基金项目“语料库驱动的隐性、显性知识接口研究”（13YYB006）的阶段成果，由扬州大学出版基金资助。由于本人才疏学浅，书中疏漏之处在所难免，敬请各位专家和同行、各位老师和同学批评指正！

陆 军

2021年8月

<b>第 1 章 绪论</b> .....	1
1.1 引言 .....	1
1.2 语料库研究概略 .....	2
1.3 型式与构式 .....	3
1.4 语料库调查的系统推进性与所谓的“重复罗列现象” .....	7
1.5 语料库参照的教学设计与基于语料库的教学设计研究 .....	9
1.6 语料库语言学与(外语)学科教学方向的教育硕士培养 .....	13
<b>第 2 章 词语搭配概念体系</b> .....	15
2.1 引言 .....	15
2.2 词语搭配 .....	15
2.3 搭配范围 .....	21
2.4 搭配行为 .....	24
2.5 二语透明搭配与半透明搭配 .....	26
2.6 隐性、显性搭配知识 .....	26
2.7 语义趋向 .....	27
2.8 语义韵与语义韵冲突 .....	29
2.9 搭配、型式与构式 .....	29
<b>第 3 章 语义韵研究的理论、方法与应用</b> .....	31
3.1 引言 .....	31
3.2 语义韵理论 .....	31
3.3 语义韵研究的基本方法 .....	35
3.4 语义韵研究领域 .....	37
3.5 结语 .....	39

<b>第4章</b>	<b>二语共选型式研究的语料库技术</b> .....	41
4.1	引言 .....	41
4.2	语料库研究的关键技术与概念 .....	41
4.3	词语索引的统计抽样技术 .....	44
4.4	语料库介绍 .....	49
4.5	基于语料库的二语分析框架 .....	52
4.6	语料库检索软件 AntConc 的操作实例 .....	57
<b>第5章</b>	<b>二语近义词搭配行为与语义韵研究</b> .....	63
5.1	引言 .....	63
5.2	文献回顾 .....	64
5.3	研究方法 .....	65
5.4	近义词搭配行为与语义韵特征 .....	67
5.5	讨论 .....	74
5.6	语料库调查结论与英语近义词教学设计 .....	75
<b>第6章</b>	<b>短语学视角下的二语词语知识研究</b> .....	83
6.1	引言 .....	83
6.2	文献回顾 .....	83
6.3	研究设计 .....	86
6.4	词语使用的短语学特征 .....	87
6.5	讨论 .....	94
6.6	结论与语义趋向知识教学设计 .....	98
<b>第7章</b>	<b>共选视阈下的二语使役态学习研究</b> .....	103
7.1	引言 .....	103
7.2	文献综述 .....	103
7.3	研究设计 .....	105
7.4	使役态的共选特征 .....	106
7.5	讨论 .....	112
7.6	结论与英语使役态教学设计 .....	114
<b>第8章</b>	<b>英语写作中定冠词 THE 的共选特征研究</b> .....	121
8.1	引言 .....	121
8.2	文献综述 .....	121
8.3	研究设计 .....	124

8.4	定冠词 THE 的使用特征 .....	126
8.5	讨论 .....	128
8.6	结论与英语定冠词 THE 教学设计 .....	131
<b>第 9 章</b>	<b>二语写作中的 (be+) V+<i>-ed</i> 型式研究 .....</b>	<b>139</b>
9.1	引言 .....	139
9.2	文献回顾 .....	140
9.3	研究设计 .....	141
9.4	(Be+) V+ <i>-ed</i> 型式的使用特征 .....	142
9.5	讨论 .....	147
9.6	结论与 (be+) V+ <i>-ed</i> 型式的教学设计探讨 .....	149
<b>第 10 章</b>	<b>同题作文词语型式与写作教学研究 .....</b>	<b>164</b>
10.1	引言 .....	164
10.2	研究方法 .....	165
10.3	动词型搭配的分布特征 .....	166
10.4	分析与讨论 .....	169
10.5	结论与基于语料库的写作教学设计 .....	171
<b>第 11 章</b>	<b>基于语料库的写作教学与评估研究 .....</b>	<b>179</b>
11.1	引言 .....	179
11.2	研究设计 .....	180
11.3	分析与讨论 .....	182
11.4	结论与基于语料库的写作评估设计 .....	186
<b>附录 1</b>	<b>CLEC 中定冠词 THE 误用示例 .....</b>	<b>195</b>
<b>附录 2</b>	<b>(Be)+V-<i>ed</i> 使用失误实例(从 CLEC 中提取) .....</b>	<b>199</b>
	<b>参考文献 .....</b>	<b>203</b>
	<b>索引 .....</b>	<b>216</b>

## 绪 论

### 1.1 引言

当下,人类已迈入大数据时代。大数据给我们的学习、工作和生活带来了前所未有的影响,诸多领域正面临着革命性挑战,语言教学与研究领域也不例外。作为语言大数据的典型代表,电子语料库的规模已从第一代百万词次的语料库(如 LOB)发展到现在的亿万词次级别,在科学研究和语言教学领域发挥着空前的作用。本书主要开展基于语料库的二语词语共选型式调查与教学设计研究。具体而言,是以基于语料库的二语词语共选型式研究为基础,探讨如何利用语料库研究发现系统评估学习者困难和分析教学需求、指导并开展相关教学设计(即基于语料库的教学设计)。其直接目的在于促进语料库语言学研究成果向教学实践转换,打通语料库语言学研究与实践相结合的“最后一公里”,解决外语词汇和语法教学的若干瓶颈问题。

作为全书的开篇之章,绪论部分首先简要介绍语料库语言学的发展现状,然后着重讨论语料库语言学和外语教学研究所面临的热点或前沿问题:一是语料库语言学的词语共选“型式”(pattern)与认知语言学的“构式”(construction)有何区别联系?二是如何认识语料库语言学词语共选型式研究的系统性与简单重复罗列?三是如何区分“参照语料库的教学设计”与“基于语料库的教学设计研究”?四是学习者语料库语言学如何能够为(外语)学科教学方向的教育硕士研究生培养服务?在逐一回答上述问题的同时,还明确了本书的理论立场,介绍了本书的篇章布局,区分了相关研究方法,阐述了本书的主要编写宗旨以及研究成果的转化路径。

## 1.2 语料库研究概略

21 世纪以来,大数据技术应用日益广泛,发展势头空前迅猛。语料库语言学作为语言大数据的重要研究领域,其发展更是如火如荼。从 20 世纪末开始开展基于单语语料库(以英语本族语为主)的词语行为或搭配行为研究以来(如 Sinclair 1991, 1996; Stubbs 1996; Kjellmer 1987, 1991; Tognini-Bonelli 2001 等),先后出现了以二语学习者语言为主要内容的中介语对比分析研究(如 Bahns 1993; Granger 1998a, b; Tono 2004; Borin & Prutz 2004; 桂诗春、杨惠中 2003; 卫乃兴 2002a, 2006; 濮建忠 2003; 李文中 2005; 陆军、卫乃兴 2014, 2017; 陆军 2012, 2018 等)、基于语料库的学术语言研究(如 Hyland 1998, 2000, 2005; Lei 2012, 2016; Lei & Liu 2016; Lei & Sheng 2016; 李晶洁、卫乃兴 2010; 卫乃兴 2016 等)、对比语料库语言学研究(Xiao 2007; Granger & Paquot 2008; 卫乃兴 2011b; 卫乃兴、陆军 2014)、语料库翻译学(如 Baker 1993; 王克非 2006, 2012; 王克非、黄立波 2011; 胡开宝 2012 等)、多模态语料库语言学(如 Adolphs & Carter 2013; 参见刘剑 2017 等)等诸多研究分支。

其中,基于语料库的二语学习者语言研究尤为独树一帜。随着首个大规模“国际学习者英语语料库”(International Corpus of Learner English, ICLE)(Granger 1998a)的开发以及中介语对比分析方法(Contrastive Interlanguage Analysis, CIA)(*ibid.* 1998b)的提出,国内外基于语料库的二语学习者语言研究蓬勃发展。其中,就词语行为方面的调查研究而言,已涉及词语搭配、类联接(colligation)、语义趋向(semantic preference)、语义韵(semantic prosody)等多个方面(参见 Howarth 1998; Nesselhauf 2003, 2005; Tono 2004; Borin & Prutz 2004; Paquot 2013; 桂诗春、杨惠中 2003; 卫乃兴 2002a, 2006; 濮建忠 2003; 娄宝翠 2004; 李文中 2005; 卫乃兴 2006; 陆军、卫乃兴 2014)。与此同时,研究路径还从经典的 CIA 研究拓展到基于语料库的元话语分析和基于用法的二语研究等方面(参见卫乃兴、陆军 2018)。经过 30 多年的发展,学习者语料库语言学研究显示出其独特的优势。大量真实语言使用数据为揭示二语典型使用特征或语言规律提供了客观的证据,大幅度提高了研究信度和效度;研究结果可直接应用于教学实践,对促进二语教学具有很大潜势。不过,作为新兴研究领域,其概念体系解读、研究方法应用、成果推介和应用前景等方面也难免遇到困难或挑战。为此,下文以“型式”(pattern) vs “构式”(construction)等概念、语料库调查研究的系统性 vs 与所谓的“重复罗列现象”、语料库参照的研究方法(corpus-illuminated/illustrated/referenced approach) vs 基于语料库的方法(corpus-based /

driven approach) 以及语料库语言学与学科英语教育硕士培养等方面的问题与挑战进行梳理。

### 1.3 型式与构式

学术交流中常常出现把基于语料库的词语共选型式研究视为构式研究,或是在构式研究中使用型式进行描述的倾向。这些倾向对相关概念和理论体系的理解会有一定影响,对相关研究成果的评估和推介也会有一定妨碍。语料库语言学的“型式”与认知语言学研究的“构式”在描述对象上存在交叉,但具有不同的学科属性和概念内涵,需要加以区分。首先需要指出的是,本书所说的语料库语言学是以英国语言学家约翰·辛克莱(John Sinclair)为主要代表的 Sinclairian Corpus Linguistics (参见 Morley & Partington 2009),以词语共选(lexical co-selection)为指导思想,以真实语言使用中的共选型式(co-selection pattern)为主要描述对象。这与把语料库语言学纯粹当作研究方法的观点不同(参见 McEnergy & Hardie 2012)。那么何为共选型式?它与普通语言描述中所使用的 pattern 概念有何差异呢?辛克莱(Sinclair 1991:3)曾指出,“Three major areas of language patterning, besides grammar, could not be comprehensively presented in a dictionary, even an innovative one like Cobuild. These are collocations, semantics, and pragmatics.”。这说明,辛克莱(Sinclair 1991)把词语搭配与传统的语法、语义和语用同等视为语言构型(language patterning)的四个主要领域;型式很重要,但是难以归纳和描述。在观察并概括大量语料库数据的基础上,辛克莱(Sinclair 1996)发现:交际中的词语总是倾向于与属于特定语法范畴、带有特定语义特征的词语共现,表达特定的态度意义,实现特定的交际目的。其中,词语与词语共选构成词语搭配,词语与语法范畴共选构成类联接,词语与语义集共选构成语义趋向,整个短语单位与态度意义共选构成语义韵。这四个范畴的共选分别构成相应的共选型式,与辛克莱(Sinclair 1991:3)所提及的四个构型领域趋于对应,它们共同构成扩展意义单位(extended units of meaning)。由此可见,所谓词语共选型式,并不是游离在传统语言描述体系(词汇、语法、语义、语用)之外看似相对独立的、零散分布的词语组合,而是趋向于把词汇、语法、语义和语用整合起来共同构成短语单位(Stewart 2010),因此,词语共选型式都是短语单位的。

约翰·辛克莱(John Sinclair)及其团队一直围绕共选型式开展研究(如 Louw 1993; Sinclair 1991, 1996, 1998, 2003, 2004a, 2007; Hunston & Francis 2000; Partington 1998; Partington et al. 2013; Tognini-Bonelli 2001; 卫乃兴 2001, 2011a, 2011b 等)。Pattern 这一术语甚至成为相关研究成果的标志。例

如,帕廷顿(Partington 1998)和帕廷顿(Partington 2013)这两部专著的标题分别为 *Patterns and Meanings: Using Corpora for English Language Research and Teaching* 和 *Patterns and Meanings in Discourse*,都在突显型式与意义的密不可分。霍斯顿、弗朗西斯(Hunston & Francis 2000)主要从类联接入手考察共选型式,并以 *Pattern Grammar* 为专著主标题。她们在书中指出,与“型式语法”在概念上关系最为密切的先驱研究可以追溯到阿尔伯特·悉尼·霍恩比(Alfred Sydney Hornby)于1954年出版的 *A Guide to Patterns and Usage in English* (ibid.: 3)。霍恩比(Hornby 1954: vi)对 pattern 有如下描述:“When [the learner] learns the meanings of the adjective he should also learn its patterns: ‘anxious about his son’s health’ ‘anxious for news’ ‘anxious (= eager) to start’.”。霍斯顿和弗朗西斯(Hunston & Francis 2000)指出,“尽管(霍恩比关于 anxious 的型式描述)看似把型式和意义割裂开来,但(……)已经很隐晦地把两者联系起来。”(ibid.: 4)霍恩比(Hornby 1954: vi)所说的形容词 *anxious* 的 pattern 其实就是其周围的词汇语法环境(environment)(ibid.: 32),而霍斯顿和弗朗西斯(Hunston & Francis)的型式语法则比词语共现抽象层级更高的共选型式,如 *it V n to-inf v-link N to-inf* 和 *it v-link ADJ to-inf* 等。由此可见,语料库语言学的共选型式研究总是与意义描述密不可分。这一点可由辛克莱(Sinclair 1996)提出的扩展意义单位模型(the model of extended units of meaning)以及辛克莱(Sinclair 2007)提出的意义移变单位(meaning-shift unit)思想进一步说明。意义移变思想指出:“共选成分的变化引起周围语义(ambient meaning)之改变,改变到一定程度会产生新的扩展意义单位”(ibid.: 3)。简言之,共选型式其实是形式、意义和功能的统一体,是短语单位的。

语料库数据除了用于共选型式研究以外,还大量用于构式研究。<sup>①</sup>“构式”这一术语由英语中的 construction 翻译而来;而 construction 这一概念可追溯到布卢姆菲尔姆(Bloomfield 1933)的《语言论》(不过,本书中更倾向于表示语言结构)。他根据直接成分分析法区分出“向心结构”(endocentric construction,如 *poor John*)和“离心结构”(exocentric construction,如 *John ran*)两类。此外,他还谈及施事-动作结构(actor-action construction,如 *John ran*)、关系-轴心结构(relation-axis construction,如 *beside John*)等。由此可见,布卢姆菲尔姆(Bloomfield)所谓的 construction 是大于单词层级的语言实体,或是词组级的、小句级的,甚至是句子级的,属句法层面的概念。与之相比,认知语言学中的“构式”则是结构与意义之结合体,并非纯句法的概念。例如,莱考夫(Lakoff 1987: 467)所使用的语法

① 以下两小节引自卫乃兴、陆军(2018)。

构式(grammatical construction)就已经体现了形式和意义配对的思想。戈德堡(Goldberg 1995, 2006)将构式明确界定为“形式-意义配对”(form-meaning pairing),且形式或功能的某一方面不能通过构成要素严格预测(not strictly predictable)。根据戈德堡(Goldberg 1995, 2003, 2006),构式的范围很广,分属不同层级:语素、单词、句法框架、习语、隐喻、稀有用法、被动式等。奥斯特曼(Ostman 2005)甚至还提出了语篇构式(discourse construction)的概念。

基于用法的语言学习研究流派(usage-based approach)认为,语言知识来源于交际中的语言使用,语言学习本质上就是在交际中获取语言结构和规则,即获取构式(Bybee, 2010; Goldberg 1995, 2006);二语学习则是在二语输入的基础上获取或归纳相应的构式(Ellis & Wulff 2015: 75)。这些观点或路径与拜比(Bybee)和兰加克(Langacker)等人的贡献分不开。拜比(Bybee 1985, 2010)强调了频率对语言使用 and 发展的影响。兰加克(Langacker 1987)则具体地指出,语言结构每使用一次都会得到一定程度的巩固(entrenchment)。

与布卢姆菲尔德(Bloomfield 1933)对 construction 的描述相比,认知语言学中的构式似乎更加强调名物化后缀“tion”(即所谓的“构式化”)对该概念的贡献。换言之,在体现“式”(structure, pattern)的同时,特别强调“形式-意义配对”或“获取”“构建”相关结构或规则知识,即看重“构”这一动态认知过程。诚然,个体接触语言现象是一个从无到有、从少到多的发展接触过程;所接触的对象包括语素、单词、句法框架、习语、隐喻、稀有用法和被动式等多个方面。相应地,他们自然会逐步把这些结构(或广义的型式)与相关意义结合起来记忆、存储,形成所谓的构式(Tomasello 2003, 2008)。构式在语言能力发展的不同阶段所起的作用并不相同。例如,在初步接触阶段,学习者可能会更多地借助语素、单词、句法框架处理相关语言现象;而在语言使用的熟练阶段,则更多地倾向于把上述相关构式组合起来使用,构成更为复杂的构式。由此可见,构式自然包括了从语素到语法框架、再到习语甚至复杂层次更高的结构,反映了人脑认知、处理语言的复杂特征。不过,它们在语言使用中并不是同等起作用的,因此并不都直接反映交际中的语言使用特征。构式的典型性不得而知,需要通过具体的认知心理实验方可揭示。但可以肯定的是,不同层级的构式并不是平均分布的。根据辛克莱(Sinclair 1991)的成语原则可以推测,在本族语交际中,类似于半构筑预制结构或短语的复杂构式趋于起主导作用,如戈德堡(Goldberg 1995, 2006)等所考察的 Causative constructions 等。

以上论述表明,典型的共选型式也颇受构式研究关注,也可采用基于语料库频数统计分析的研究方法。尽管如此,两个概念之间存在着本质差异。首先,它们所涵盖的范围不同,典型型式都可能是构式研究的考察对象,而构式中有很多

层级却不属于典型的共选型式(如词素、词、稀有用法等)。其次,型式反映了(词语)结伴使用,说明了意义的可预测性(You shall know a word by the company it keeps, Firth 1957),属于同一型式的不同词语共享部分意义(Hunston & Francis 2000)。基于语料库的型式调查发现,可以根据搭配词、类联接、语义趋向预测节点词的意思;同时,由于节点词与其词汇语法环境共选、共同实现意义,因此可以根据搭配词、类联接和语义趋向预测语义韵。这些都反映了交际过程中的意义实现特征,即话语社团反复使用某一型式、构筑特定的语义韵、形成特定的短语意义单位。这种短语单位具有唯一的语义韵(陆军、吴茜 2019)。与之相比,构式研究则强调多义性、不可预测性(参见 Goldberg 1995: 31-32, 2006: 5)。构式的多义性说明了构式需要和其他构式共现才能够表达确定的意义,而不能通过构式内部的构成要素进行预测(即所谓不可预测性)。因此,单个构式的意义需要形式—意义配对的个体认知过程形成。

按照上述界定,扩展意义单位是典型的共选型式,但未必是构式。根据成语原则(Sinclair 1991),这些扩展意义单位应该是话语社团共享的语言知识。因此,如果说构式认为话语成员认知存储语素、单词、句法框架、习语、隐喻、稀有用法和被动式等各个方面,则更倾向于反映人类对语言构成要素的认知,主要是记忆和规则训练所获得的具体知识,很多为显性知识;而共选型式则倾向于反映话语社团对词语搭配、类联接、语义趋向和语义韵型式的集体认知,是语言使用中通过统计学习(statistical learning)规律所形成的知识(R. Ellis 2004; N. Ellis 1994; Ellis & Frey 2009; Ellis, Frey & Jalknean 2009),多为隐性知识。

就一语习得研究而言,学习者有大量机会接触语言使用、发生统计学习,这使得观察不同层级构式要素的分布特征及其发展变化特别有意义。但就二语习得而言,词素、单词、语法结构、隐喻、稀有用法、被动式等构式有很多可以通过相应的双语对应结构进行认知和习得,也是传统的二语教学研究的关切。然而,由于构式具有多义性,不具有可预测性(Goldberg 1995: 31-32, 2006: 5),二语学习者即使能够获得大量目标语构式,也未必知道其具体意义和用法、如何对他们进行组合。当二语学习者在表达中缺乏目标构式的组合知识时,倾向于依靠双语对应词联系选择目标语对应词和语法结构进行表达(Jarvis & Pavlenko 2008; 桂诗春 2004)。换言之,二语学习者所缺乏的主要知识并不是目标语构式的形式—意义配对知识,更主要的是如何把这些构式要素进行组合,形成符合目标语表达习惯的词语序列(参见 Howarth 1998: 36)。相应地,二语学习者语言研究首先需要调查二语表达在组合关系上的特征,即二语词语组合、二语词语与语法组合、二语词语与搭配词语义集组合,以及这些组合与所表达的态度意义的组合是否符合目标语表达习惯。这些都是学习者语料库语言学共选型式研究所关