

前言

词汇作为语言三大要素(语音、词汇和语法)中的基本建筑材料(桂诗春 2006),在语言中的地位不言而喻。英国语言学家Wilkins(1972: 111)关于词汇与语法关系的经典论断“While without grammar little can be conveyed, without vocabulary nothing can be conveyed”几乎成为论及词汇的必备语录。Clark(1993: 1)也在论述习得中的词汇时开宗明义,断言是词汇成就了语言(“Words make a language”)。没有词汇就没有音、形和句法结构。因此,词汇是语言的中心,词汇亦已取代句法成为Noam Chomsky最简方案理论框架的主要特征。语言的习得在很大程度上就是词汇的习得(Gass & Selinker 2001)。

词汇习得虽然重要,但其发展并非一帆风顺。一语习得领域的词汇研究姑且不论,自结构主义语言学兴起至20世纪80年代,二语习得领域都未将词汇习得纳入其研究视野,无论是在语言教学理论还是教材大纲中,都难以找寻词汇教、学、测的踪影(Mitchell & Myles 2004; O'Dell 1997)。虽然一直有学者为词汇应有的地位在努力,比如英国的Harold E. Palmer、美国的Michael West等曾在20世纪初发起以词汇学习为基础的口语和阅读学习的词汇控制运动(vocabulary control movement),但在结构主义和句法优先占主导的环境里,词汇一直难以突破“灰姑娘”(Meara 1980)的身份。词汇学习自身的确不甚系统,亦

不像句法规则那样有章可循,但规则离开了词汇便无法表述(结构主义)或生成(生成语法);若真要用好句法规则,实现圆满交际,那么到底需要掌握多少词汇也是需要认真考量的,这些因素也就决定了词汇习得的不可忽略性(Milton 2009)。

现代意义上的词汇习得研究始于Meara(1980)在*Language Teaching & Linguistics: Abstracts*杂志上发表的论文“Vocabulary acquisition: A neglected aspect of language learning”。他在文中总结了关于词汇控制、记忆、存储、语义测试等的实验及研究现状,呼吁学者重视词汇习得研究。此后,越来越多的学者开始关注并投入词汇习得研究中。《二语习得研究》(*Studies in Second Language Acquisition*)杂志分别于1987年和1999年策划了词汇研究专刊,标志着词汇研究被冷落状态的终结和被聚焦状态的开始,但词汇研究发展成为二语习得的研究热点和核心领域则是近30年的事(Coady & Huckin 1996; Nation 2001; Schmitt 2010)。二语词汇习得研究呈现出发展迅速、学科交叉、视角多元的喜人气象,既有词汇习得的经典话题研究,如词汇的广度与深度、附带习得与有意学习、学习的策略与测量、词汇的习得与语言技能等;又有传统的词汇学习与习得的延伸研究,如词汇的记忆与提取、识别与辨认、存储与联想、维持与磨损等;更有从不同视角探究词汇习得的研究,如心理加工与词汇习得、语料库与词汇习得、认知—神经与词汇习得、社会文化视角下的词汇习得等。近年来,丰富多样的研究技术和实验手段使词汇习得研究蓬勃发展,最近10余年间发表的词汇研究成果约占过去100年总量的30%以上(de Bot 2015)。这其中不乏非常活跃高产且极具影响力的学者,比如曾组建二语词汇习得研究小组的Paul Meara教授。他也是呼吁关注词汇研究的第一人,几十年来一直带领着团队成员致力于二语词汇习得各个领域的研究。他本人也在拓展词汇研究的疆域中发挥着先锋和领路人的作用,是二语词汇研究的领军人物之一。Paul Nation教授在二语词汇教学领域的地位举足轻重,尤其是在二语词汇教与学的测量和策略方面。他的*Teaching*

and Learning Vocabulary (1990) 和 *Learning Vocabulary in Another Language* (2001) 两部著作被视为词汇教学的必读之作。Norbert Schmitt 教授作为二语词汇习得研究领域发表研究成果最多的一位学者，其于2008年在 *Language Teaching Research* 上发表的综述是该期刊20年来引用率最高的文章。活跃在二语词汇习得研究领域的学者还有聚焦词汇测试的 Averil Coxhead、James Milton 和 James Read 等，侧重词汇附带习得的 Batia Laufer、Jan H. Hulstijn 和 Stuart Webb 等，关注语言技能与词汇习得的 Marjorie Bingham Wesche、Ronard Carter、Tahereh Sima Paribakht 和 Tom Cobb 等，关注语块研究的 Alison Wray、Frank Boers 和 Sylviane Granger 等，在词汇加工、联想、磨蚀等领域享有盛名的 Bert Weltens、Brent Wolter、David Singleton、Joe Barcroft、Nan Jiang 和 Tess Fitzpatrick 等。

词汇习得研究除却一向热门的词汇教学主题之外，近年来也出现了其他更丰富、更聚焦的研究话题，比如 Lengyel & Navracics (2007)、Fitzpatrick & Barfield (2009) 和 Tokowicz (2015) 对二语学习者词加工的讨论，Hoey (2005) 提出的词启动理论，Bogaards & Laufer (2004) 和 Nation & Webb (2011) 对词汇研究的全面介绍，Milton (2009)、Schmitt (2015) 和 Meara & Miralpeix (2016) 对测量二语词汇的研究方法和工具的推介，以及 Szudarski (2017) 对如何借助语料库进行词汇研究的专门介绍。

国内的二语词汇习得研究者紧跟国际同行的脚步，从关注中国英语学习者的词汇量开始，拓展到词汇学习策略分析与策略培训研究、基于语料库的词汇使用特征研究、词汇掌握与语言技能的关系研究、词汇知识的发展研究、词汇习得的认知机制研究、词汇理论的提出与验证研究、词汇习得的心理加工研究等。二语词汇习得的研究队伍几乎是中国二语习得研究中体量最大的队伍，其中不乏如鲍贵、陈士法、范琳、卢植、马广惠、倪传斌、孙蓝、张淑静、张文忠、赵晨等资深学者，更有如陆军、王

新朋、吴诗玉、张北镇、张会平、郑咏滢、钟志英等后起之秀。研究对象从小学生到硕士研究生，几乎涵盖了所有阶段的英语学习者；研究成果丰富且不乏力作。遗憾的是，迄今为止，虽然学者众多、成果累累，有关词汇习得的著述却不多见。除却以相关博士论文为基础出版的著作和马广惠(2016)论述英语词汇教学研究的专著，鲜见其他。对于国内广大的二语词汇习得研究者，尤其是对二语词汇习得感兴趣的初入门者来说，他们迫切需要全面了解二语词汇习得领域的研究进展和未来走向，以帮助自己更好地进行学术定位。本书正是基于这样的背景和需求而编写。

二语习得通常指母语之外的其他语言的习得，但现有文献基本以英语为研究对象，以法语、德语、西班牙语、韩语、日语、俄语等为研究对象的文献比例相对较低。近年来有关将汉语作为第二语言的来华学习者的汉语词汇习得研究数量有所上升。考虑到汉外学界与英外学界的教学现状和作者的学术背景，本书以英语作为第二语言(外语)的词汇习得研究为主体，不多涉及其他语种的词汇习得。

全书由七章组成。第一章概述与词有关的术语。首先从词的相关概念入手，介绍了什么是词、什么是搭配、什么是词库；进而厘清了词习得的四组概念：词汇的广度与深度，接受性与产出性词汇，附带习得与有意学习，词汇维持、自动化与磨蚀；最后从词汇知识和词汇能力两个方面介绍了词汇测量的基本概念与常用范式。第二章梳理词汇习得的相关理论。首先阐述了关于词汇知识的三种不同观点：成分观、发展观和隐喻观；然后列举了九种解释词汇学习的理论或假设；最后总结了四个有关词汇发展的理论模型。第三至七章为实证研究，分别从词汇知识的习得，课堂环境下的教习得，词汇学习不同策略，基于语料库的词汇习得，以及认知语言学、心理加工、动态系统和社会文化视角等方面介绍了二语词汇习得从词形到词义、从单个词到多词、从知识到技能、从教学到策略、从习得到磨蚀、从使用到加工等多视角、全方位的研究成果和发展趋势。此外，每章均提出了本领域的研究建议，推荐了相关文献，实证研究部分还总结了

适用的研究方法。正文最后提供了参考文献，如果读者对某一领域感兴趣，可以进行深入的拓展性阅读。另外，考虑到本书的读者可能对二语词汇习得研究中的某些概念不太熟悉，我们还添加了对应的英文术语，供读者参考。最后三个附录分别为“常用词汇习得教与学及测量的网站资源”“二语词汇习得研究的主要学者及研究领域”“常用二语词汇习得的测量工具及分类”。除此之外，本书还提供了主要术语索引，方便读者查询和使用。

本书适合语言学和教育学等专业的研究生阅读，尤其适合从事中高等教育语言教师和研究者作为词汇教学与研究的案头读本。

本书的出版需要特别感谢外语教学与研究出版社的支持。解碧琰编辑为本书的出版付出了很多心血。她在本书初稿完成之际通读了全书，并给予了非常中肯和详细的建议，在本书的结构调整和文字校读方面多次与作者协调磋商，力争为读者提供最友好的入门书籍。华南理工大学/华南师范大学的陈艳艳博士对本书理论部分有重大贡献；中原工学院的段士平老师对本书第3.3小节和第3.4小节的搭配和词块部分有重大贡献；华南师范大学的俞建耀老师对词汇学习策略部分有重大贡献。笔者的研究生方南和王文艺、李丹、李阳妹、林欣欣等小伙伴一起，收集并完成了三个附录的基本内容。部分2017级、2018级和2019级研究生参与了全书参考文献格式的校对。陈艳艳、吕文璐、刘嘉玲、陈卓等不仅是本书的第一批阅读者，更是费心费力地复校、制图并提出修改建议，替笔者承担了很多琐碎且费时的工作。没有她们的全力参与和倾情付出，本书难有可能在此呈现。必须要坦白的是，即便我们再努力，由于能力所及、精力所限、研究成果日新月异，这样的一本书很难做到一应俱全、赏心悦目。倘若挂一漏万，错责在己。

张 萍

苏州大学外国语学院

开展任何一项研究之前都首先要对相关概念进行界定。词汇习得研究涉及的概念当然是词。“词”听起来人人皆懂，但越是众所周知的概念越难以有统一的、人人皆认可的定义。词到底是什么？所学/习得之词存储在何处？习得词需要了解和运用词的哪些知识？如何测量已学词汇的掌握程度？本章第一节厘清词、搭配和词库等基本概念；第二节介绍词汇习得领域的基本术语；第三节解释词汇知识和词汇能力的测量概念。本章每一节都会介绍今后的研究趋势，并提出建议，推荐相关文献，供感兴趣的读者拓展阅读。

1.1 词的概念

本节首先介绍关于如何界定“词”的若干纷争，然后对词与词结合形成的两词及两词以上的搭配的定义进行简单梳理，最后解读若干词集合在一起形成的词库。本节并非局限于对词、搭配、词库基本定义或概念的铺陈，更多的是呈现与之相关的研究成果或悬而未决的争议。因此，本节最后介绍了相应的研究趋势，并提出建议。

1.1.1 什么是词?

如何界定“词”一直是一个令人困扰的问题。古希腊和古罗马学者对词的功能和表意争论不止，中外学者对“什么是词”也众说纷纭，莫衷一是。因此，给词下一个工作定义不是一件容易的事情，更何况对词的理解是二语词汇教学和习得的基本问题，涉及词汇习得研究的理论基础。那么，到底什么是词？

亚里士多德认为，词应包含四个方面的知识：真实世界的存在、个人印象或概念、词的发音形式和词的书写形式。现代语言学之父索绪尔 (Saussure 1916) 使用能指 (signifier) 和所指 (signified) 两个概念分别表示词的形式和意义。Vygotsky (1962) 根据这样的二分法思维，从社会文化视角提出语言中每个词所代表的意义都是语言和思想相结合的产物。结构主义遵循索绪尔的理论思路，认为词是音和义之间任意的、约定俗成的联结，是具有形式和意义双重特征的语言单位。是否学得或习得一个词，最简单、最基本的判断方法就是确认是否了解这个词的形式和意义，即是否建立了词形和词义之间的联系，也叫形—义匹配 (form-meaning mapping)。例如，我们在日常言语交际中看到的是词的书写形式，听到的是词的语音形式，理解的是词的意义，但词形和词义不可能都是一一对应的关系，即便是“词”，在英语里也有如 word、lexicon、lexis、vocabulary¹ 等不同的表达。单个词也有基本义和引申义，还有同形异义、一词多义、同音异义，以及词与词组合之后形成的不同语义透明度、不同语境下词义的变化等。每个单词就像钻石一样有很多棱面，难以通过简单的形—义匹配标准来解释。

Palmer (1921) 首次区分了在语境中能识别的词和在言语、写作时能自如使用的词。在语境中的词还包括搭配和固定短语等。他用 polylogs 指

1 按 Read (2000) 等的理解，word 通常指语言中的词单位；lexicon 指语言中习得词汇的存储单位，即词库；lexis 指含有词位和词目的词汇层或某一特定范围的词汇；vocabulary 指语言中词的总和。

代搭配,用miologs表示词缀知识,即词素(morpheme)。从词汇学研究的角度来看,一个词要包含以下几个方面。一是语素成分。当我们看到unbelievable这个词时,要知道它的词根、前后缀及其位置,比如un-是对believable的否定,而不是把-able当作unbelieve的后缀。二是语法特征。要了解这个词是名词还是动词,若是名词,则名词是否可数;若是动词,则动词是否及物,以及A词是否可以和B词搭配等。三是语义知识。一个词不可能孤立存在于语言和语言使用中,世间万物有千丝万缕的联系,投射到语言的词汇上就会有语义关联的表达,如上位词(hypernym)(动物)、同位词(coordination)(狮子、老虎、羚羊、斑马等)、对立词(antonym)(寒冷—炎热、年轻—年老),这些构成一个词的语义网络。四是语用知识。实际使用词汇时需要在不同语境下采用不同的词或语体,有时要用委婉语,有时要用正式体。

Hanks(2013)认为,一个词不是简单地由字母串组成的意义体,词的概念体系应包含词的类符(type)、形符(token)、词目(lemma)、词位(lexeme)、专有名词、多词短语等。类符指发音或拼写形式不同于其他形符的词符,例如work、teacher、student、worker、teaching等都属于不同的类符。形符则指书面语或口语中出现的每一个词符,不论其形式是否相同、是否重复出现,都记作一个形符。词位指词的语音、拼写和形态信息,是最小的词汇单位;而词目则指该词具有的语义和句法信息。Levelt(1989)将词所包含的词位和词目信息浓缩成图1.1,成为后续学者广为接受的词项内部结构模型。Jiang(2000)在论述二语词汇习得的三个阶段时就采用了该模型,并一针见血地指出二语学习者的词汇习得很难达到本族语者的水平,原因在于二语学习者习得的二语词汇介入了过多的母语词汇语义和句法信息,他们的二语词汇更像是穿着二语词汇的外衣(词位信息),裹着一语词汇的内核(词目信息)(图1.2)。

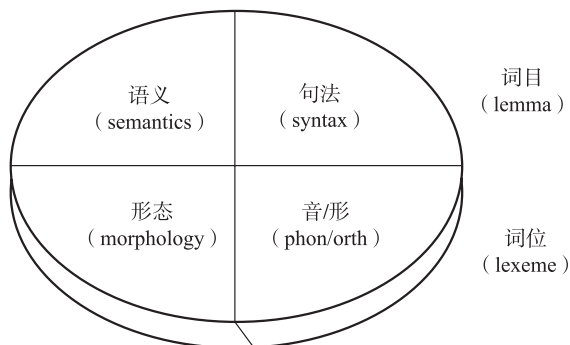


图 1.1 词项的内部结构 (Levelt 1989)

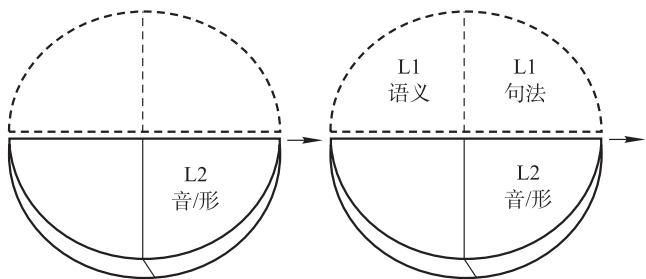


图 1.2 二语词汇的固化 (Jiang 2000)

词的分类也有众多标准。有功能词和内容词之分、低频词和高频词之别，有书面语和口语词之分、基本词和专门词之别，有单个词和多词之分、个词和词族之别。Carroll *et al.* (1971) 完全以词形为标准，认为一个符合拼写规则的词串就是一个词，不需要区分衍生或屈折等形态。曾有一个专门筛选心理语言学实验用词的网站列出一个包含 30 多个特征的清单，以帮助用户选取实验用词，也就是说，几乎每个词都承载了众多的词汇特征，比如拼写、音素 (phoneme)、音节、词频、具体性 (concreteness)、意象性 (imageability)、词性 (part of speech)、语境状态、发音变异性 (variability)、重音模式以及屈折 (inflection) 规则性等。

由此可见,对“什么是词”这一问题无法用一句话来回答。学者们退而求其次,转向求证“怎么才能算是掌握了一个词”,希望从这个角度来曲线构建词的概念体系。Cronbach(1942)也许是现代语言学史上最早将词的知识划分为不同成分的学者。他认为了解一个词需要具备五个方面的知识:词的定义义(generalization)、词的使用义(application)、词的多种义(breadth of meaning)、词的精准义(precision of meaning)和词的支配力(availability)。Richards(1976)提出衡量是否真正掌握一个词的八条标准,即必须了解这个词的词频、语域(register)、句法特点、语义特点、屈折变化、联想义、概念义以及一词多义,但未提及词的语音、拼写形式以及与他词的共现特征。Wallace(1982)紧随其后,在补充完善其标准的基础上提出,二语学习者如果希望能像一语者那样去“认知”所学词汇,则必须能够:1)识别词汇的书面语和口语形式,2)随时想起该词,3)与相关概念或实体进行联想,4)准确使用语法形式,5)发音无误,6)拼写正确,7)搭配得当,8)语体使用得体,9)知晓词汇的内涵与联想义。如果考虑到二语学习者已有的母语词汇知识,则还要区分是从一语还是二语的角度来界定一个二语词,因为有些词可能更适合从一种语言的词形或词义角度去界定,从而更有效地区分两种语言间的词汇概念。Nation(1990)针对前人提出的标准结构松散、逻辑不强等问题,进一步细化了Richards(1976)的假设,认为二语学习者必须知晓词的八个方面的知识:口语形式、书面语形式、语法形式、搭配、频率、语域、概念和联想义。目前学界最为认可的是经Nation(2001)修改过的词汇知识框架,从形式、意义和使用三个方面详细列举了词汇知识涉及的不同成分,例如对词形的掌握包括对词的发音、拼写和词结构的识别和运用,对词义的掌握包括了解词形和词义的匹配、词的概念和所指义以及词的联想义,对词的使用的掌握包括了解该词的语法功能、搭配共现、使用频率及语域限制等。能否正确、得体、有效地使用二语词汇,体现的就是学习者的二语词汇能力。

1.1.2 什么是搭配?

词是语言的基石,但词不能孤立地存在。无论是依赖词与词之间语义连贯的衔接关系,还是依靠词与词之间句法构建的共现关系,语言中的每一个词都不可避免地与其他词产生共生相伴的机会。从构词的角度来看, collocation 中的 col 源自 com-, 表 together、with 等“一起”的意思; -loc- 表 to place or put “放置”的意思,因此, collocation 本身就表示“经常一起出现的词”的意思。传统语言学从句法视角分析语言中词与词之间的关系,而关注社会文化影响的欧洲语言学派重视语境对语言和词汇的作用,认为词的意义需要通过与之相伴的搭配词 (collocate) 得以更好地体现,搭配本身也是词的意义的一部分 (Firth 1957)。

最早的研究是 Palmer (1933) 系统收集的英语习语和短语动词报告。他将“词语搭配”定义为“由两个或两个以上的词组成,必须作为一个整体来学习,而不能从其组成部分任意拼凑起意义的词组”(1933: i)。他从语法角度将搭配分为两类:一类是由实词与实词,如动词+名词组成的搭配;另一类是由实词与虚词,如动词+介词组成的搭配。后续研究逐渐扩大了搭配的范围,又因为语言学与各个领域的联系日益密切,各种各样的词语搭配开始被统一纳入短语研究,并拥有了不同的术语:心理语言学称为程式语 (formulaic language) (Wray 2002); 功能语言学称为词汇短语 (lexical phrase) (Nattinger & DeCarrico 1992); 语料库语言学称为多词单位 (multi-word unit)、词串 (lexical bundle)、意义拓展单位 (extended unit of meaning)、N元语块 (n-gram) (Biber & Conrad 1999; Sinclair 1996; Stubbs 2003)。多词单位的范围从谚语到词串再到两词搭配,目前还没有完全统一的选择和分类标准。Howarth (1998) 将搭配分为词组合和语法组合关系,认为搭配是一个从自由搭配、限制搭配到比喻和成语的连续统 (continuum)。Cowie (2002) 按照词语的长度 (length)、语义的透明度 (transparency) 和组合 (composition) 的固定程度等将短语分为自由搭配、有限搭配、修辞性成语、纯成语、程式语等子类。Moon (2002) 根

据机构化 (institutionalization)、固着性 (fixedness) 和不可组合性 (non-compositionality) 将多词短语分为复合词型 (如 crystal ball、navy blue)、动词短语 (如 stretch out、tick off)、常用谚语 (如 have an axe to grind、rain cats and dogs)、固定短语 (如 of course、you know) 以及预制语块 (如 the point is、that reminds me) 等类型。

从注重语言的句法结构和功能到关注语言的词汇习得和加工, 语言习得的研究重点随着研究手段的更新而变化。语料库因其特有的语言使用特性, 能提供极为丰富和鲜活的搭配语料, 而搭配又必须是基于频率、依赖语境的实际使用, 因此, 语料库的发展极大地激发了学者们进行搭配研究的热情, 从而使界定搭配及其他多字词语有了更为量化的指标。Nattinger & DeCarrico (1992) 对“搭配”的定义是: (句子) 水平维度上与某个节点 (node) 词左右共现的跨距 (span) 组, 其垂直维度上的跨距词可以因语境而被其他词替代, 构成不同的组合 (见图 1.3)。

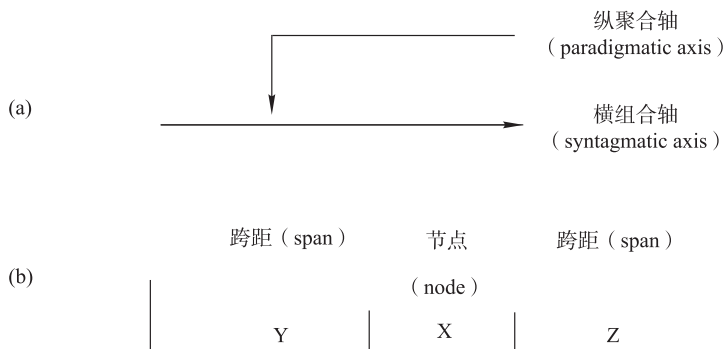


图 1.3 “搭配”的定义 (Nattinger & DeCarrico 1992)

当该节点词与跨距词在文本中互为期待 (mutual expectancy) 的共现频率高于概率值时, 就构成了一个搭配。节点词和跨距词的共现频率越高, 该搭配关系就越稳固, 当共现频率高到两者总是固定相伴出现时, 搭配就变为谚语, 两者的横—纵关系完全固化至无变异可能, 词义的选择空间也被

大大压缩直至“失义”，或者叫做语义不透明 (semantic opacity)。kick the bucket就是常被用来举例的典型谚语。如果两者中的跨距词有多种共现关系，那么该搭配的语义保存就会更完整，共现的变异性也会更高，即该节点词可以有更多与之相伴的跨距词，构成语义完全透明到半透明的搭配关系。例如高频动词find，可以与具体名词work、job、key、book等构成语义非常透明的自由搭配，也可以与抽象名词happiness、comfort、guilt、pleasure等构成限制性的半透明搭配，直至与feet构成语义不透明的谚语find one's feet (适应了新环境、能独自行动)。

搭配的划分从基于语法规则的空位填充到语义不透明的词汇组合，同样符合Sinclair (1991) 的开放性原则 (open-choice principle) 和习语原则 (idiom principle)。Hakuta (1974) 区分了固定语块 (unvarying chunk or routine) 和预制语块 (prefabricated pattern)。前者是固定的词相伴，如what's that; 后者是固定的结构，但部分词可以被替换或移动，如wanna+v。她认为语块不仅可以事先记忆，也可以随语境根据句法规则临时组合。但在生成语言学家眼里，这些语块已游离至基于规则的语言链的边缘，属于规则之外的创造性使用 (如Clark 1974; Krashen & Scarcella 1978)。

1.1.3 什么是词库?

词库 (lexicon) 在二语词汇习得领域是一个较新的概念，源自心理语言学。Aitchison于1987年出版了第一本正式讨论词库的专著 *Words in the Mind: An Introduction to the Mental Lexicon*，引发了学界对词库或心理词库的研究热潮。词库，顾名思义，就是人类存储词的仓库 (human word-store) (Aitchison 1994: viii)，或叫心理词库 (mental lexicon)，是一种语言中所有词汇的集合 (Ježek 2016)，是使用或理解语言时可以随时从中提取的词的仓库 (Clark 1993)。最早使用mental lexicon的是Ann Treisman (Coltheart *et al.* 2001)。她在1961年的博士论文里提出，一个人所学的词汇可以像心理词典那样，存储在词的仓库里。

词库可以是单语词库，如Clark(1993)专门讨论了儿童一语习得中的词库；也可以是双语甚至多语词库，如Schreuder & Weltens(1993)和Pavlenko(2009)的研究。

当然，心理词库肉眼不可见。为便于理解，语言学家们提出了很多比喻，如McCarthy(1990)的百科全书说(encyclopaedia)、Aitchison(1994)的蜘蛛网说(web)、Murphy(2010)的同义词典说(thesaurus)、Brown(2006)的电脑网络说(network)等。心理词典(mental dictionary)是最常用的类比说法。事实上，词库与词典除了存储的共性之外，更多的是差异性。词典通常以字母顺序排列，试图全面描述一种(单语词典)或两种(双语词典)语言的词汇集合或对应的纸质和电子介质，是具体的物质存在。词库则是抽象概念，是由各种词群组合关系，如语素族、语义网、句法特征等组成的词汇网络。词典也可以模拟心理词库的结构，基于词与词之间的语义关系编制，如McArthur(1997)编写的《朗文多功能分类词典》(*Longman Lexicon of Contemporary English*)就打破了传统词典和分类词典的拘囿，按语义关系将同义、近义和语义相关的词编排在一起。词典编纂当然可以遵循不同的原则，但没有一部词典可以列出某种语言中所有的词汇和使用规则。另外，词典编写的内容时效久远、静态不变、规整简洁，而词库则随时更新、动态变化、结构多样。因此，学习者习得的词汇量会远小于词典所列词的数量，存储在词库中的某些词汇也会因一些创造性使用而不被词典所罗列。此外，二语词库中的词需要经过多频次的输入、明确的感知、有意或附带的学习和练习才有可能被固定在词库的中心地带。如果将心理词库想象成一个同心圆，那么随着语言水平和使用频率的变化，词的存储状态会在词库中不断变化，越来越熟悉的词会逐渐向词库的中心靠拢，而不常用的低频词则会被排挤到词库的外延地带，因此，词库的结构其实并不像Wolter(2001)描绘的那样，不同熟悉程度的词之间是相互隔离、不可“穿越”的(见图1.4)，而应该像张萍(2013)所类比的那样，具有动态发展性(见图1.5)。

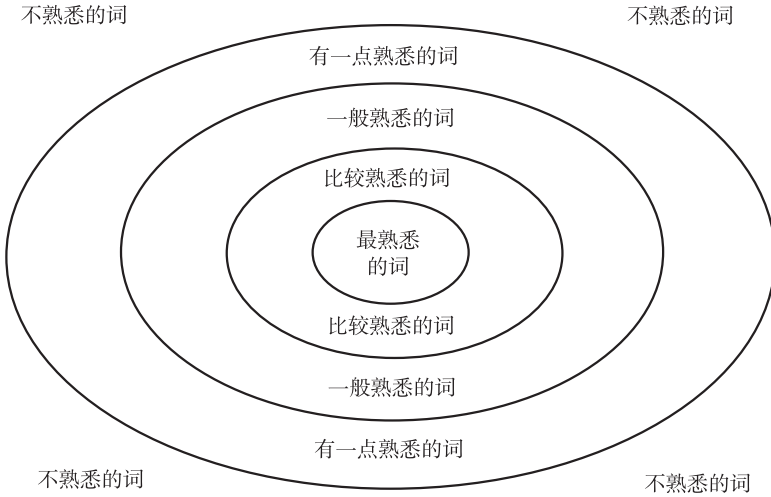


图 1.4 二语心理词库的知识模型 (Wolter 2001)

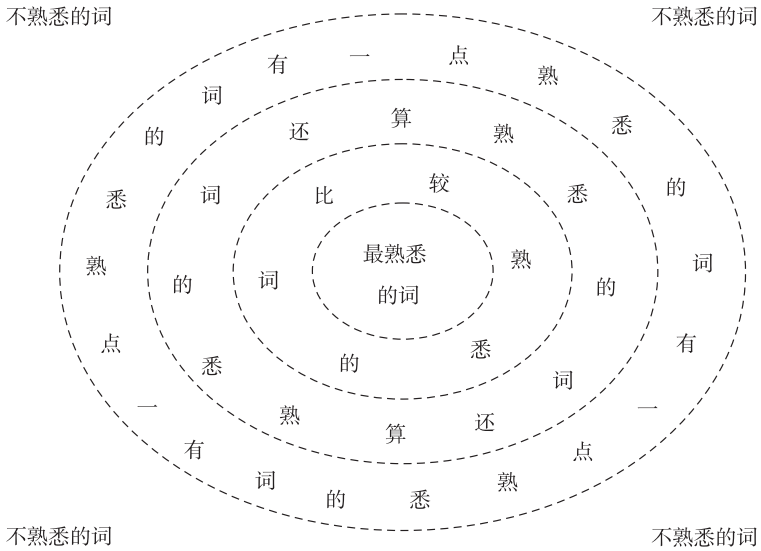


图 1.5 二语心理词库的组织模型 (张萍 2013)

心理语言学关于词库的研究主要集中在心理词库的存储方式和表征模式上。学者们对一语者的心理词库组织有相对一致的认识,即随着年龄的增长和语言水平的提高,不论一语词库的词类如何(如词性、具体性、词频等),其词汇组织逐渐由语音联结向语义关系转变、由句法搭配关系向纵聚合语义转变,也就是通常所说的音—义(phonological-semantic,简称P-S)迁移和横—纵(syntagmatic-paradigmatic,简称S-P)迁移(Entwisle 1966; Ervin 1961)。最初有关二语词库的研究认为,二语词库应该和一语词库的发展路径一样,经历从音到义、从横组合到纵聚合的发展过程(Singleton 1999; Zareva 2007)。随着研究的深入,学者们逐渐发现一语系统的存在对二语词汇学习产生了不可避免的影响,最终造成二语词库独有的组织和发展模式,其语义网络远不如一语词库那样丰富和紧密,而且非语义化联结关系的比例要远高于一语词库。最为明显的是,不论学习者的语言水平如何,二语词库中纵聚合的语义联结都要大大强于词与词之间的共现关系,呈现出音—义混合、纵先于横的二语词库结构。一语词库研究将纵聚合联结的增强作为一语词汇能力发展的指标,而二语词库研究则将横组合知识的增加作为二语词汇能力发展的标志(Wolter 2002; 张萍 2010)。二语词库的组织不同于—语词库,更易受词类效应影响。名词的语义联结要比动词的语义联结更紧密和稳定,随着语言水平的提高,名词的语义关系更容易建构。形容词比动词的语义联结更丰富,发展更快,但都集中在纵聚合的拓展而非横组合的共现上(Navracscics 2007)。在语义联结和纵聚合关系上,高频词和具体词也都比低频词和抽象词表现出更好的关联度(张萍 2011)。

有相关模型对二语词库内部词汇间的存储模式进行了解释。最早的模型是Weinreich(1953)提出的混合—独立模型。通过对词汇联想(word association)反应结果进行分析,他发现二语学习者的词汇存储系统至少有三种形式:一是一语与二语各自独立,二是两者共享同一系统,三是二

语词汇借助一语通达语义概念。Potter *et al.* (1984) 将混合—独立模型改造为并列表征模型 (parallel representation model), 认为在语言水平较低的阶段, 二语学习者需要借助一语与二语概念产生联系, 而在词形层面, 两种语言则分别表征。随着语言水平的提高, 二语学习者可以直接通达二语词汇的概念层。

为了更好地解释语言水平对心理词库表征的影响, Kroll & Stewart (1994) 修正了之前提出的层级模型 (hierarchical model), 明确了一语系统影响下, 二语词库中词汇与语义概念联结强度变化的关系 (图 1.6): 二语词汇与概念的联结强度弱于一语词汇与概念的联结强度, 且二语的词库容量远小于一语的词库容量。针对概念特征在二语词库中的作用, de Groot *et al.* (1994) 分析了一系列翻译—识别任务中具体词和抽象词的实验结果, 提出了特征分布模型 (distributed feature model, 简称DFM) (图 1.7)。该模型将词汇的语义概念分解为若干概念元素, 认为表达相同意义的一语和二语词汇并不完全共享所有的概念元素, 且具体词和抽象词共享的概念元素比例也不一样, 具体词共享两种语言间概念元素的比例要高于抽象词。共享的概念元素比例越高, 该词汇就越容易被习得, 存储也就越稳定。

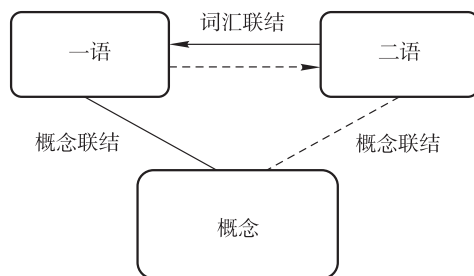


图 1.6 修正层级模型 (Kroll & Stewart 1994)

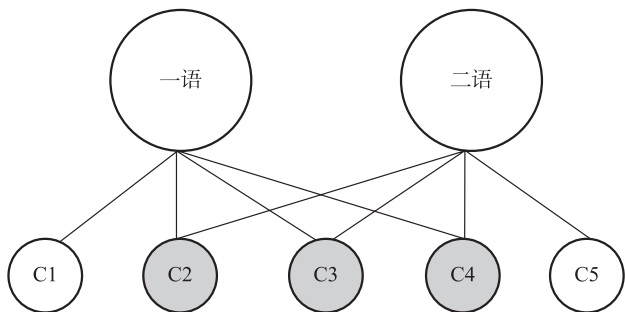


图 1.7 特征分布模型 (de Groot *et al.* 1994)

de Groot *et al.* (1994) 的模型基于对具体词和抽象词的一语—二语间双向翻译任务结果的分析, 后续学者又采用词汇联想任务进一步完善特征分布模型, 并提出了共享分布式不对称模型 (shared distributed asymmetrical model) (图 1.8) (Dong *et al.* 2005)。该模型将词汇的语义概念分解为一语、二语和共享三类概念元素, 认为一语和二语翻译对等词 (translation equivalent) 享有共同的概念元素, 但一语与共享概念元素的联结强度要高于二语与共享概念元素的联结强度。一语与其独有概念元素、二语与其独有概念元素的联结均要强于各自与对方独有概念元素的联结, 形成分布式、不对称的词汇—概念联结模式。相比于以往提出的解释模型, 共享分布式不对称模型将二语独有的概念元素分离出来, 强调二语词汇不能与一语词汇意义重叠的不对称性。但该模型仍未能涉及二语词库中词汇句法知识的表征特点。通过对中国英语学习者词汇联想任务中具体词和抽象词的联想反应结果进行分析, 张萍 (2011) 提出了共享概念与独有知识分布模型 (图 1.9)。中国学生的联想反应结果表明, 无论是具体词还是抽象词, 他们都能针对所给刺激词联想出更多与世界、事物、概念等有共同认知的百科知识或原型概念, 而其中很多知识都归属于语义或纵聚合大类反应类型。如果两种语言特有的词汇共现或搭配知识的共享元素较少, 那么反映在联想结果中就是横组合关系的数量很少, 而且会出现较多一语独有的搭配。

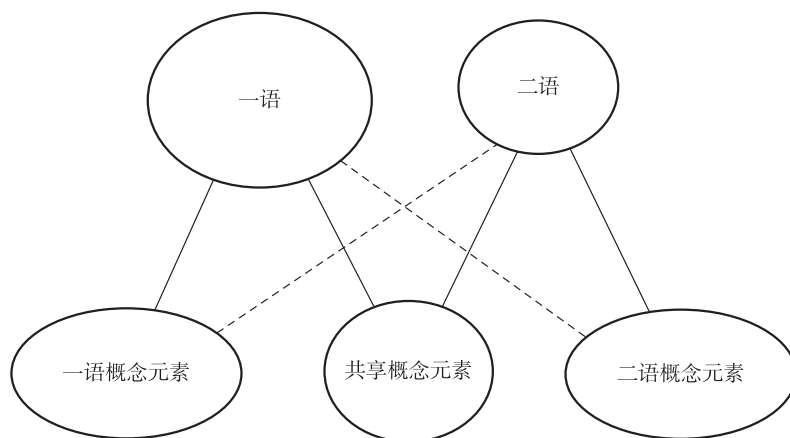


图1.8 共享分布式不对称模型 (Dong et al. 2005)

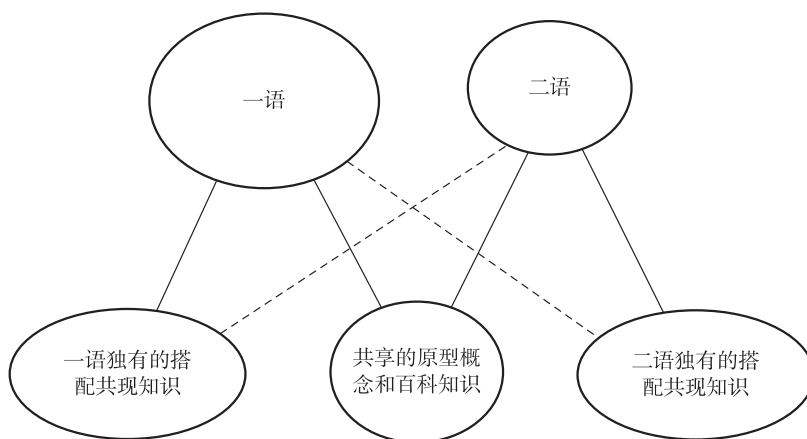


图1.9 共享概念与独有知识分布模型(张萍 2011)

1.1.4 研究趋势与建议

1) 词的定义

已有研究对词进行的分解基于词汇知识的成分观。该观点认为，掌握一个词需要了解该词的不同成分和不同维度的知识，这样的认知有助于词

汇教学内容的细化和词汇测量的实施,因此,也是目前广泛使用的词汇知识理论体系。但也有学者认为,如果从每个维度去衡量学习者对词汇知识的掌握程度,则会遇到不少实际问题:首先,在测量学习者的词汇知识习得程度时,很难一次性顾及所有维度;其次,不同维度的词汇知识并非总是按照同一速度、以同样的轨迹发展;最后,测量所得的词汇习得结果并不一定是想要观察的某一维度知识,而可能包含了其他方面的词汇知识能力。因此,在词汇知识的数量和质量、线性和发展、测量的信度和效度等方面一直存在争议和分歧,这是今后研究仍需关注的焦点之一。

从生成语言学视角来看,词是信息模块的载体,包含与之相关的所有语言信息(如语音、词形、语素、句法、语义),并在我们的大脑中构成一个信息储藏库,无论选词造句还是说话写字,该词的所有信息都通过句法层面投射出来。每个单词好像一块拼图,携带着自己独特的信息,并与其他单词相衔接,组成合法又合体的句子。词在生成语言学家眼里是一个个信息丰富、等待词同伴来完美速配的词子,但如果落地到具体的词汇教学、词汇测量和词汇习得中,这些表征信息是预先存储还是实时加工、心理现实性如何、生成语言学如何解释心理加工现象等,都是未来研究面临的挑战。

2) 搭配研究

如何理解和界定“搭配”,至今仍有争议。基于短语学对搭配进行的解读强调明确搭配的词性差异(实词与实词组合还是实词与虚词组合)、语义透明和黏着程度;基于频率对搭配进行的区分更多强调单个词和组合词共现的频率及强度,忽略对搭配语义的区分。对搭配的解释恐怕不能从单一视角出发,否则只能是“盲人摸象”,结果支离破碎而无整体图景。

对“搭配”这一术语的认同也是未来基础性研究需要解决的问题。有学者将搭配等同于多词单位、预制语块、词串等;有学者严格区分搭配与短语、谚语等,只将两词组合称为搭配。不同的研究基于不同的认知,学者们按照自己的标准选择和设计实验材料,导致所得实验结果难以在同一

层面上进行对比和观照，这也是目前很多研究面临的共同问题。未来的搭配研究还是需要术语的内涵进行相对统一的认定。

采用什么样的研究方法进行搭配研究亦不容忽视。有基于语料库的统计和对比研究，有语料库驱动的搭配研究，有从心理加工角度解读存储和记忆特征的研究，有对比语料库和联想常模的同配程度研究。采用何种研究方法以及如何结合多种研究方法，都是现在和将来很长一段时期内搭配研究的热点和焦点。本书还将在第3.3小节展开介绍。

3) 心理词库

近年来有关二语词库的研究方兴未艾。尤其是关于一语词库和二语词库是各自独立还是融为一体，存在各种说法和证据。坚持分离观的一派有两个理论基础：一种观点是模块假说(modular hypothesis)，认为人的大脑由相互独立、功能各异、结构分明的不同模块组成(Fodor 1983)，其中一个模块专司语言，因此一语习得的词库是内模块的(intramodular)，而儿童期之后学得的二语是外模块的(extramodular)，后学语言中的词汇显然不属于先前的语言体系，应属于另一套存储系统；另一种观点认为，两种语言在词形上存在很大的差异，特别是当一语词库中没有二语词汇的词形知识时，就不可能再将新的二语词汇存储到一语词库中(Bybee 1988; Stemberger & MacWhinney 1988)。

支持融合论的一派态度也很坚决，但对一语词库和二语词库融为一体的程度有不同看法。一种观点认为一语词库和二语词库没有完全独立或完全融合，因为两个词库共同存在于同一个大脑，而二语学习者又能独立使用两种语言，但两个词库之间存在高度联动和动态互动(Herdina & Jessner 2002)。例证来自心理语言学的词形/词义判断任务(morphological/semantic judgment task)。二语词形会激活两个词库中词形相似的词汇，尤其是形态相近的近形词(Dijkstra 2001)，语境线索也无法抑制多义词或同形异义词的启动。

最新研究对词库的关注热度不减，从语言哲学领域到神经语言领

域，对词库的再定义和词库的心理现实性展开了持续的讨论和探索。争议聚焦在以下两个方面。

争论之一是词库存储的内容。有学者认为词库只存储语言变异的乖僻词条 (Chomsky 1995; Pinker 1998); 有学者坚持词库就是经过长期记忆的词、短语、谚语等各类词储库 (Jackendoff 2002; Ullman *et al.* 2005); 更有学者提出词库不仅存储词的语言学意义, 还存储个体对世界认识的百科全书知识 (encyclopedic knowledge) (Begby 2016)。

争议之二是多库论、单库论和无库论的假设。Fodor (1983) 提出模块式思维, 并对语言加工认知模块进行切分, 认为大脑中应有分别存储概念、语音表征、输出声音和物体表征的库。关于脑成像、失语症和老年痴呆症的研究结果支持分库式表征。Ullman (2007) 和 Teichmann *et al.* (2012) 都发现语音 (中—后颞上回) 和拼写 (颞下回) 是分开的, 语义 (前—中颞叶层)、句法 (前额皮质层) 和词汇 (颞叶) 是分离的。支持单库论的学者多受联通论 (connectionism) 的影响, 认为人的大脑由神经元及连接突触 (synapse) 组成, 习得的知识不是存储在不同功能的模块中, 而是分散在输入、隐藏和输出的层级神经网络中, 经过平行加工后得以提取 (McClelland & Cleeremans 2009)。交互激活模型 (interactive activation model, 简称 IAM) (McClelland & Rumelhart 1981)、TRACE 系列模型 (McClelland & Elman 1986) 等都是联通论视角下的单库论解释模型。联通论和基于使用理论 (Usage-based Theory) 的一些支持者甚至提出根本就没有心理词库这种说法, 认为语言的所有知识元素, 从语音、拼写到句法、语义直至语用信息都是高度融合的有机体, 在同一个网络里彼此相连, 随时互动, 且随语境而不断动态变化 (de Sousa & Gabriel 2015)。Pulvermüller (1999) 对大脑左右侧的皮质活动进行分析后, 发现左侧主词形, 右侧主词义, 发音和声学表征都在布洛卡区和韦尼克区的外侧裂周区附近。语言的学习牵涉到大脑的整体参与, 而大脑左右侧的功能分布在某种程度上也是对索绪尔形—义二分法的一种支持, 换句话说, 形—义

匹配是一个即时在线、不断强化的自动化过程，而非预先存储在一个子虚乌有的心理词库中等待被激活或被提取。

迄今为止，有关词库的谜团仍未解开。人类现有的能力还不能全面深刻地了解和认识大脑的认知和加工机制，因此，词库到底应该被看作知识存储还是界面接口，实体存在还是理论装置，均有待后续研究对词库的功能作进一步的探索和解惑。

推荐文献

Aitchison, J. 1994. *Words in the Mind: An Introduction to the Mental Lexicon* (2nd ed.). Oxford: Blackwell.

Chomsky, N. 1995. *The Minimalist Program*. Cambridge: Massachusetts Institute of Technology Press.

Hanks, P. 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge: Massachusetts Institute of Technology Press.

Nation, I. S. P. 2013. *Learning Vocabulary in Another Language* (2nd ed.). Cambridge: Cambridge University Press.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

1.2 词习得的概念

词汇习得就是学习和使用词汇知识的过程，首先是掌握一定量的词汇，即词汇广度/宽度 (breadth/width)，其次是加深对词汇知识的了解和使用，即词汇深度 (depth)。也有学者从不同角度提出，对词汇知识的