

第二语言词汇习得的语料库研究方法^①

邢红兵

(北京语言大学 汉语水平考试中心, 北京 100083)

[摘要] 本文首先分析了目前词汇习得研究中存在的对频率、偏误、习得过程、句法语义关系、目的语和母语的关系等方面认识的局限, 提出建立用于二语词汇习得研究的词汇知识库, 并在此基础上进行词汇知识习得研究。我们建议第二语言词汇习得应该围绕词汇知识系统性、发展性的原则, 采用因素分析思路进行对比分析。

[关键词] 第二语言; 词汇知识; 词汇习得; 语料库

[中图分类号] H195.1 [文献标识码] A [文章编号] 1003-7365(2012)02-0077-09

零、引言

随着第二语言习得研究的逐步深入, 语料库资源在第二语言习得研究中占据越来越重要的地位。究其原因, 可以概括如下: (1) 语料库资源建设的逐步丰富和发展。由于语料库技术的发展, 用于语言习得研究的目的语和中介语语料库资源也越来越丰富, 使用也越来越容易, 相关的研究成果也越来越丰富。根据顾晓波(2010)对2009年以前的9年间国内研究者研究方法所做的统计, 基于语料库的中介语研究成果大幅度上升, 但是受语料库的限制, 历时研究较少, 共时研究比较多; 从研究方法上看, 实证研究占绝对优势; 从发展趋势上看, 国内基于语料库的中介语研究已经由理论方法引介及单纯的错误分析转向中介语对比分析的研究。(2) 语言学习理论的发展。比如联结主义理论的建立, 对语言习得研究产生了重要的影响。作为联结主义的核心浮现特征揭示了语言知识的获得和语言材料之间的本质联系, 浮现特征的核心原则是语言规则不是先天固有的, 而是在语言获得过程中表现出自然浮现特征, 这种浮现过程可以基于对大量输入语料的统计学习而获得。(Rumelhart & McClelland 1986; Bates et al 1998; MacWhinney 1999) 联结主义理论认为语言习得的过程实际上是学习者通过真实语言材料获取语言知识并储存于心理词典中的过程。(3) 语言对比在外语教学研究中的作用。随着语料库资源的进一步丰富, 语言对比研究也逐渐深入。从上个世纪九十年代开始, 随着中介语语料库的建设及其加工过程的深入, 中介语对比方法使二语习得的对比研究能够深入到中介语系统中。(4) 词汇习得研究越来越受到重视。文秋芳、王立非(2004)认为上个世纪80年代末90年代初, 人们开始重视词汇的习得, 成果逐渐增多, 主要的研究领域包括二语作文中词汇知识的变化、词汇丰富性的研究、被动词汇和积极词汇之间的关系、阅读中词汇知识的广度与深度的关系的研究等。这些研究在方法上深受认知心理学的影响, 大多数研究采用量化法。

① [基金项目] 本研究得到教育部人文社会科学重点研究基地重大项目“留学生汉语词汇习得的计算机模拟研究”(项目编号: 08JJD740063) 和教育部人文社会科学研究一般项目“面向对外汉语教学的语言知识表达系统研究”(项目编号: 09YJAZH013) 的资助。

由此可见,利用语料库资源,对二语词汇知识习得进行对比研究是目前研究的发展方向。但是如何利用语料库资源进行二语习得研究,特别是如何在理论上取得突破,还需要进一步探讨。从目前的研究现状来看,基于中介语语料库的汉语作为第二语言习得的研究主要还是从偏误分析的角度来进行的,尽管偏误分析能够在一定程度上了解留学生习得目标语的一些特点,但是,仅仅从偏误分析的角度出发,我们无法对中介语系统中知识的存储和加工过程等方面的整体情况进行深入的了解。

一、中介语语料库词汇习得研究的误区

至今,我们利用中介语语料库进行习得研究主要是采用偏误分析方法,因此,我们也把留学生错误的数量作为衡量二语习得状况的重要指标。但是就目前利用语料库进行词语习得研究的状况来看,在认识上和方法上还存在的问题。其研究误区可概括为以下六个方面。

1.1 使用频率和习得效果呈正相关

一般认为,词语在中介语语料库中使用次数越高,该词语的习得就越好。在中介语语料库中,我们能够准确地统计出二语学习者使用某个词语的频率。因此,我们常常使用中介语的频率来解释习得情况。但是,是不是使用越多就学得越好呢?这其中存在一个泛化和简单化的问题。比如,“们”在汉语中除了出现在“我们”、“你们”、“他们”等词语中构词以外,其他情况下使用很少。但是在“中介语语料库”中,“们”除了同样出现在“他们”等代词中以外,还大量被使用,使用次数达到725次,频率是1677.6815(频率的基数是一百万,下同),而“现代汉语语料库”中的频率是1.6298,差别非常大。实际上,这些例句中的“们”很多是多余的。例如:

- ①现在,我的同学们都在北京,我在日本。
- ②所以他常常出去找这些孩子们和年青人跟他们说话。

再比如虚词“的话”,在中介语语料库中使用了338次,频率是768.2624,而这个词语在“现代汉语语料库”中的频率是38.2997,差别非常明显,但是从我们的分析来看,很多例句中的“的话”显得多余,更多的例句中,“的话”的语体色彩不合适。例如:

- ③你们要是去中国旅游的话,肯定能看见这样情况。
- ④如果有个同学生病了的话,他很关心照顾病人。

值得注意的是,常用词使用泛化和非常用词使用贫乏的情况。在中介语中某个词语的使用次数高,并不能完全代表对该词语的运用就很好。中介语语料库中词语的使用频率应该参照母语语料库的情况,使用过多可能表明出现了一定的泛化情况。在中介语语料库中存在常用词使用频率明显比现代汉语高的情况,而且还比较普遍。

1.2 正确率是衡量习得效果的直接指标

目前我们对于中介语语料库中词语使用情况的衡量标准就是错误率,一般的认识是:错误率越高,习得越差;正确率越高,习得越好;错误率应该随着学习者水平的提高而逐渐下降。如果我们采用心理学的实验设计研究方法来研究某个项目的习得,错误率是衡量习得过程的重要指标,这是因为实验的项目(称之为“因素”)以及各个项目的不同特征(称之为“因素的水平”)是经过设计的,实验是在匹配各种条件的情况下进行的。因此,实验结果中的错误率和习得情况直接相关。但是在语料库研究中,学习者的条件、使用环境、使用次数、搭配类型等因素都会影响错误率,不能简单地将中介语语料库中的错误率和习得直接联系起来。

因此,我们要将错误率和相关的因素结合起来进行具体分析。这些因素可以从两个层面来认识:(1)词语的使用情况。使用的类型越丰富,出现的错误可能越多,随之也会出现较高的错误率。(2)错误的类型。在二语习得过程中,词语的使用可能会出现不同类型的错误,每种错误在每个阶段的错误率可能会不一样,因此,错误率要相对于不同的错误类型。如许尔茜(2011)对留学生“V着”的习得进行了研究,译文按照“V着”的结果体、进行体和未完成体进行分析,发现错误率分别是25.90%、40.39%和26.47%。按照这个结果,我们一般认为进行体最难,未完成体较容易,但是对比现代汉语的使用情况,译文发现未完成体的比例在现代汉语中(16.53%)和在中介语中(2.83%)的比例相差很大,这表明学生在中介语语料库中使用上有明显的欠缺,因此仅仅凭借错误率来断定未完成体习得好这样的结论是不准确的。

1.3 词汇习得过程主要是词语意义的习得过程

长期以来,我们注重词义系统的教学,主张词义是词汇教学的核心,比如教学过程中注重同义词辨析的问题,相关的研究也是注重词义系统的对比和分析。虽然现有的研究已经注意到将同义词的习得范围拓展到易混淆词范围之内来进行分析,也将词汇习得研究的视角扩大到一组词,而不是传统意义的同义词范围,比如动词“看”和“检查、见、见到、看到、看见、了解、欣赏、游览、展示、观、见、看见、瞧、上、斜视”等词语都可能混用(蔡北国2010),但是,即便如此,我们还必须意识到,二语词汇习得过程中更重要的影响因素远不局限于此。我们认为,母语相关词汇及其使用情况、目的语目标词与其相关词汇的关系、目的语与母语词语等的对应关系、各类词语在目的语的搭配词语及搭配频率等因素都会影响词汇的习得过程,而且是第二语言词汇习得的关键因素。因此,可以说,第二语言词汇习得实际上是一个从意义到用法的实现过程,其学习的难点在于意义到用法的转变。

1.4 二语词语习得是一个渐进的过程

从目前的研究来看,虽然有些研究者提出了二语词汇习得的阶段性(Jiang 2000; 邢红兵2009),也有人提出了关于理解性词汇(receptive vocabulary)和产出性词汇(productive vocabulary)的区分,如Nation(1991)就曾提出了一个明确的定义,严格区分了这两类词汇的内涵,但实质上这两类词汇知识习得的本质的区分在于词语词汇知识的获得和运用。严格来说,中介语语料库中出现的词汇尽管都是产出性词汇,但是,并不能说明能够产出一个词就意味已经习得这个词了,因为词汇知识的获得是动态发展的过程,词汇知识需要在二语的使用中逐渐去丰富,并逐渐接近母语者的词汇知识水平。一个词在二语学习者习得的各个阶段中,可能出现不同的正确率,这正是词汇知识不断获得并调整的动态过程。

因此,词汇习得不仅仅是一个简单的会与不会的过程,而是一个词汇知识逐渐丰富的过程,也是一个逐渐接近目的语词汇知识的过程,这个过程是漫长的,需要长时间的积累。

1.5 词语使用中句法、语义和功能是各自独立的过程

对外汉语教学研究重视语法教学,词汇教学及其相关研究相对薄弱,形成了语法规则为主导的教学模式,淡化了词汇知识的获得,特别是词汇知识系统的认识及其教学应用。而词汇知识本身包含词语的句法信息和语义信息,并且两种信息并不是截然分开的。比如,在中介语语料库中,留学生在“高兴”的时候,常常会出现诸如“这事儿真高兴”和“高兴的心里”这样的错误搭配,其原因在于留学生学习了“高兴的事儿”和“心里高兴”这样的搭配关系以后依据语义关系而进行了句法功能类推。张银丹(2011)发现,留学生在习得汉语形容词的时候,存在句法功能和

语义关系相互作用的情况,学习者掌握好句法、语义关系需要一个很长的过程。这表明词汇习得实际上是句法、语义知识共同作用而又相互影响的过程。因此,从词汇知识角度来看,词汇知识是句法关系和语义关系相互作用而形成的一个复杂的知识体系。

1.6 母语词汇和目标语词汇存在对应关系

在我们使用的汉语教材中,词语的对应翻译方式在词汇教学中占有很重要的地位,表现为对课文中新出现的词语,采用英语或者学习者母语进行翻译,这种翻译方法为二语学习者提供了方便,但同时也导致了一些问题。周琳(2007)曾对教材中的同译词语在中介语语料库中的使用情况进行了调查,发现同译词语的注释方式增加了留学生出现偏误的可能,特别是意义关系较远的同译词语更容易出错。究其原因,主要是学习者在获得意义的同时,也将母语相关的使用知识直接迁移到二语词汇中。众所周知,不同语言的词汇,尽管我们认为是对译关系,但是用法上的差异要远远大于意义上的差异。汪慧慧(2010)考察发现,汉语和英语常用动词所带的宾语及其频率上,两种语言的对应动词的宾语搭配很少有完全对应的,也很少有完全不对应的,更多的情况是一种不完全对应的关系。请见下图:

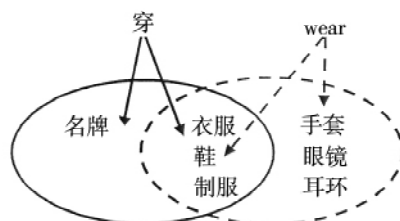


图1 汉语动词和英语动词带宾语对比示意图(引自汪慧慧2010)

从这个结果来看,第二语言学习者需要在区分两种语言之间关系的基础上建立自己的双语词汇表征体系,学习的过程实际上更是建立两个心理词典并形成各自的表征,以及建立两个词典之间的各种关系的过程。因此,学习者母语词汇知识和目的语词汇知识关系的建立,在二语词汇习得研究中就显得至关重要,二语词汇知识习得实际上就是摆脱母语对应词用法对目的语词汇用法知识的影响过程。

二、用于对比的词汇知识库建设

2.1 中介语系统与中介语语料库

王建勤(2010)认为,早期的中介语理论是在 Selinker、Corder 和 Nemser 三位学者所提出的理论主张的基础上形成的。早期的中介语系统具有如下特点:(1)中介语具有独立性;(2)中介语是可以观察到的语言系统;(3)中介语具有系统性;(4)中介语具有动态发展特征。我们认为中介语系统从中介语理论产生伊始,就具备了这样的特征。中介语语料库是在中介语理论基础建立的,是二语学习者产出的书面语或者口语语言的集合,是了解二语学习者学习过程和特点的重要途径。

作为观察中介语系统的依据,中介语语料库除了具备中介语的特点以外,还要能够体现二语词汇知识体系。因此,还应该具备如下特点:(1)发展性。中介语语料库最重要的特点就是要体现学习者目的语词汇知识的不同发展过程。二语学习者在习得过程中,由于目的语知识水平的不同,表现出明显的阶段性特点,中介语语料库应该比较全面地涵盖二语学习者学习目的语的词汇知识发展的不同阶段。因此,一个理想的中介语语料库系统,应该均衡收录处在不同水平的二

语学习者的语言表现,能够准确反映二语学习者词汇运用的真实情况。(2)分布性。中介语语料库也应该在充分考虑学习者母语背景基础上,形成代表不同母语背景学习者的语言特性。由于中介语是介于目的语和母语之间的中间状态,因此,中介语语言系统受到母语和目的语的共同影响。不同母语背景的二语学习者产生的语言的集合,代表该语言学习者的词汇习得的特点。(3)系统性。中介语语料库基本代表了中介语系统,因此具有系统性的特点。和目的语相比,中介语系统典型的表现就是正确和错误、地道和生硬、泛化和缺失、单纯和丰富并存,同时也在不断发展这样的变化过程。词汇是语料库的构成基础,通过对中介语语料库中词汇的观察,可以发展中介语语料库词汇知识。

2.2 第二语言学习者的目的语词汇知识体系

邢红兵(2009)将二语词汇知识的获得过程分为三个阶段,其中最重要的阶段就是二语学习者词汇知识从意义到用法的转变,因为二语词汇习得的关键是从意义层面向用法层面转变,而中介语语料库更多体现的是词语的实际用法,因此,基于语料库的词汇习得研究就更应该通过用法来了解学习者对词汇知识的掌握情况。而第二语言词汇知识的习得最终要达到接近目的语词语的用法的目标,就应该建立一套用于二语词汇习得的目的语词汇知识体系。下面我们以形容词“简单”为例,来分析词汇知识的构成情况。请见下图:

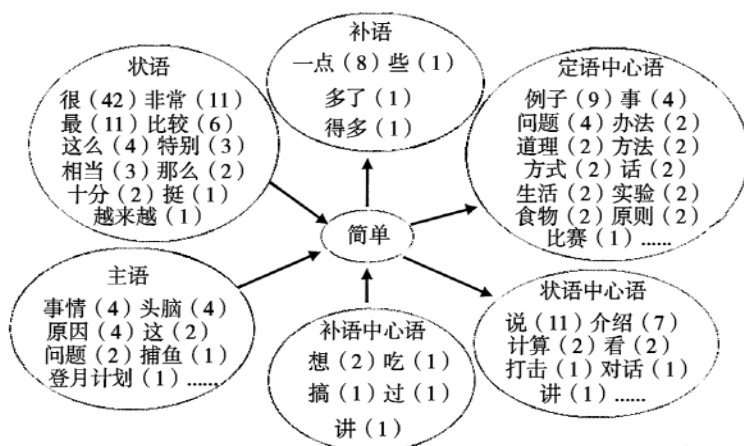


图2 “简单”的搭配知识体系示意图

从图2可见,形容词“简单”与相关的词语建立了联系,并形成了关联性的词语关系体系,包括句法搭配关系、语义关系、题元关系、聚类关系、频率知识等,这些都是词汇知识的具体表现。因此,这套词汇知识体系应该包括以下两个方面:(1)词语使用知识,即词语在真实语言中的使用情况。就目前的语料库研究成果来看,我们可以提取到词语在实际语言中的使用次数、词语的句法属性、词语之间的搭配关系、词语的语义搭配类型、词语的搭配范围等。这些属性共同构成了词汇知识的主体部分。(2)词语关系知识。比如,根据语义形成的同类词语、搭配关系形成的搭配词语、修饰关系词语等。图2中每个圆圈中的词语实际上已经形成了聚类关系,比如“简单”的全部状语形成了状语聚类关系,而“很”、“非常”、“最”等表示程度的状语又形成了一个句法语义聚类。同时,“简单”的同类词语“容易”也会有一个和“简单”具有一定共性和差异的知识体系,这两个知识体系重合部分又会使“简单”和“容易”形成了近义关系,这也属于我们所讨论的词语关系知识。

三、利用语料库进行第二语言词汇习得研究的原则

根据前面的分析,我们认为利用语料库资源进行第二语言词汇知识习得研究需要遵循以下几个方面的原则。

3.1 系统化原则

词汇知识在语料库中表现出系统性,体现在以下两个方面:(1)每个词语在目的语母语者语料库中实际使用所形成的词汇知识本身就是一个系统,目的语词汇知识系统实际上就是二语学习者词汇发展的目标。因此,二语词汇知识获得的研究首先要从词汇知识系统本身出发,考察二语词语习得情况的重要指标就是目的语词汇知识系统,单纯的使用错误率或者搭配错误率并不能全面衡量二语词汇习得的过程。(2)每个词语在词汇系统中,并不是独立存在的,而是按照各种关系聚集在一起,形成诸如反义关系、同义关系、动宾搭配、主谓搭配等关系。我们对中介语语料库的研究必须将具有各类关系的词语看成是一个个系统,词汇习得的过程就是一组关系词语的选择以及使用过程。从系统化的角度进行研究,就可以避免单纯的频率因素在二语习得研究中的局限,也可以丰富偏误分析的结果。

3.2 发展性原则

二语学习者的词汇知识是一个逐渐丰富的过程,在这个过程中,会受到目的语词汇知识和母语词汇知识的共同影响,在中介语语料库中表现为介于母语和目的语中间的状态。随着目的语水平的提高,学习者的词汇知识体系逐渐接近目的语。前面我们以形容词“简单”为例,分析了“简单”在目的语中的搭配关系及其频率构成的词汇知识体系。我们按照同样的思路对中介语语料库中的搭配知识进行了提取,以“简单”充当谓语时的主语的搭配词语为例,并根据是否重叠,形成如下图所示的交叉关系(其中实线的椭圆内为目的语的搭配词语,虚线的椭圆内为中介语的搭配词语)。

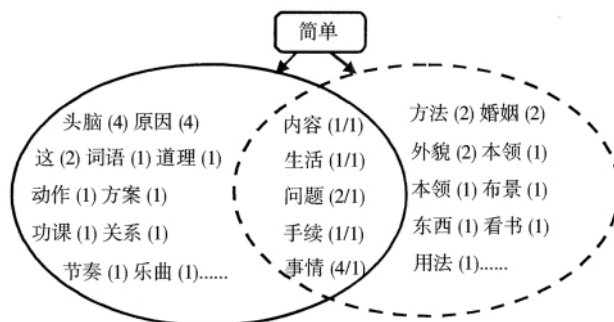


图3 中介语和目的语词汇搭配知识对比示意图

在二语词汇知识获得过程中,可以从以下几个方面来观察学习者词汇知识的变化过程:(1)使用度。从中介语语料库词语的使用来看,随着学习者目的语水平的逐步提高,词语的使用从出现泛化和欠缺逐渐转换到均衡。比如“事情”、“头脑”和“简单”形成了主谓关系,在现代汉语语料库中使用了4次,而“事情”在中介语语料库中只用了1次,“头脑”没有使用。这表明两个词语在中介语语料库中主语搭配上的使用度偏低,而上文提到的“们”、“的话”和目的语比较,明显使用度偏高。(2)丰富性。二语学习者目的语词汇知识的获得是一个从简单到丰富的渐变过程,表现在搭配关系由简单到逐渐丰富,因此丰富性是我们衡量二语学习者词汇知识的重要指标。比如在现代汉语语料库中,“简单”的主语成分有48个词语,而在中介语语料库中出现的是23个,

而且数量是随着学习者汉语水平的提高逐步增加的。因此,我们认为中介语语料库中“简单”的主语搭配词语的丰富性还不够,和目的语还有一定的差距。(3) 匹配度。匹配度是指中介语语料库中词语的使用情况和目的语使用情况的重合情况。比如上文图3中现代汉语和中介语中和“简单”搭配的主语成分中,两个语料库中的搭配并不完全匹配,只有“问题”、“事情”、“内容”、“生活”和“手续”5个词语是两个语料库中共有的搭配,因此,我们认为,中介语语料库中“简单”充当谓语时主语成分的分布情况和现代汉语主语成分的匹配情况并不十分理想,学习者还需要进一步学习,获得更多的搭配知识,进而获得更高的匹配度。

3.3 对比分析原则

对比分析理论和语料库语言学在第二语言习得研究方面的结合,产生了利用语料库资源对比中介语和目的语等的对比研究方法,其中 Granger(1996/1998/2002)提出的中介语对比方法最具代表性。Granger(2002)认为,基于学习者语料库的研究方法包括两个方面:(1) 中介语对比方法(CIA, Contrastive Interlanguage Analysis) 这种方法是对母语者(NS Native Speaker)和二语者(NNS Non-native Speaker)之间的语言进行定量和定性的研究;(2) 计算机辅助错误分析(CEA, Computer-aided Error Analysis) 这种方法是利用计算机对中介语中的错误进行标注、检索和分析。就目前的研究来看,如何将基于语料库的对比分析真正应用到词汇知识的获得研究中,还缺少较为系统的考察,研究的角度也需要进一步扩展。

我们认为,应该将中介语对比研究与语料库研究结合起来,克服第二语言词汇习得研究的误区,建立起词汇知识体系,按照一定的原则进行二语习得词汇知识获得过程的研究。这可从对比的范围和内容上来分析:(1) 对比的范围。参照 Granger(1996/1998/2002)的设想,基于语料库的中介语词汇习得研究的范围可以概括为以下几个方面:(a) 中介语(IL)和目的语(TL)的对比研究;(b) 中介语(IL)和母语(NL)的对比研究;(c) 母语(NL)和目的语(TL)的对比研究;(d) 中介语系统自身体系中不同阶段、不同母语背景等属性的对比研究。这方面的对比研究应该包括中介语系统发展的不同阶段和不同母语背景的学习者,中介语体系的对比研究等方面。(2) 对比的内容。对比分析的内容主要是以目的语词汇知识为目标,以母语的词汇知识作为参照,对中介语系统中各个母语背景、各个阶段的属性进行对照分析。对比分析要抓住使用度、丰富性和匹配度等重要环节来进行。

3.4 多因素分析原则

实验设计(experimental design)是心理学研究的最重要手段,而在实验研究中,变量的设立控制也是实验研究的关键。通过多因素的实验设计,可以分析变量的不同水平以及不同变量之间的关系。我们以舒华(1994)的研究为例,来说明因素分析的重要性(见表1)。

表1 教学手段和不同学习能力学生学习成绩表(引自舒华 1994)

	b1	b2	Σ
a1	80	78	79
a2	92	64	78
Σ	86	66	

假定有两种教学方法, a1 表示正常讲授教学, a2 表示独立学习和讨论的教学方法;有两组学生, b1 是学习能力强的学生, b2 是学习能力弱的学生。这里面已经涉及到了两个因素。如果不考虑学生的学习能力,那么结果就是两种教学方法没有差异,但是如果考虑学生的学习能力,就发现第二种教学方法对学习能力强学生有提高,但是对学习能力的学生来说,

效果比正常讲授还要差。这就是多因素分析及因素之间交互作用的典型例证。

词汇知识是一个复杂的知识系统,因此,词汇知识的相关研究就会涉及到很多不同的因素。从目前基于语料库的词汇知识体系的研究来看,有几个因素是习得研究的关键因素,主要包含以下几个方面:(1)语言材料的自身特征;(2)二语学习者的母语背景;(3)二语学习者的目的语水平。其中语言材料的自身属性最为复杂,需要我们进行认真细致的分析。从词汇习得角度来看,应该对词汇本身的属性进行细化。比如可以从动词类型和母语背景两个因素来分析词语习得过程,也可以选取心理动词和动作动词作为动词类型因素,同时选择日语、韩语和英语背景的留学生的习得情况进行分析,这样我们就可以看到不同语言背景的学习者在习得这两类动词上是否存在差别。

四、结论

总的来说,我们认为基于语料库的词语知识体系建设及其相关研究在第二语言习得中应该得到重视。第二语言习得过程是语言知识体系不断完善并最终形成系统的过程,目标语料库的属性是二语习得研究的目标参照体系,母语词汇知识对二语词汇习得有一定的影响,二语词汇习得研究的语料库方法中需要重视对比方法的运用,语言习得研究要重视语言知识体系中多种相关因素的相互作用。利用语料库进行第二语言词汇习得研究,要在对比分析的基础上,将词汇知识习得研究的视野由单纯的词汇意义扩展到词语的用法层面,研究中要将词汇知识看成是一套完整的系统,学习词汇的过程就是词汇系统中词汇知识的建立及词语关系建立的过程。每个词所具有的词汇知识就是一个系统,词汇的习得过程就是词汇系统的发展过程。要对目的语、母语和中介语系统进行词汇知识的对比,并且将学习者的阶段性、学习者自身的学习特点等反映在对比分析中。词汇习得过程涉及到很多因素,因此,进行词汇习得研究就必须采用多因素实验设计分析的方法和思路,重视各种因素之间的相互关系。

中介语语料库系统较好地反映了二语学习者的语言产出情况,但并不能完全代表二语学习者的心理过程,这主要体现在:(1)语料库偏重于语言产出,对语言理解的观察有限;(2)对二语学习者的心理机制的了解不全面,因为仅从语言输出结果并不能直接观察到学习者的词汇知识体系的特点。因此,我们还要采用实验的方法更直接地了解二语学习者词汇知识建立的心理过程及其变化情况。

参考文献:

- [1] 蔡北国. 中介语动作动词混用的调查与分析[J]. 世界汉语教学 2010 (3).
- [2] 顾晓波. 基于语料库的中介语研究述评[J]. 辽宁中医药大学学报 2010 (3).
- [3] 舒华. 心理与教育研究中的多因素实验设计[M]. 北京: 北京师范大学出版社, 1994.
- [4] 汪慧慧. 基于英汉动宾搭配对比的留学生词汇习得研究[D]. 北京语言大学硕士学位论文 2010.
- [5] 王建勤. 第二语言习得研究[M]. 北京: 商务印书馆 2010.
- [6] 文秋芳, 王立非. 二语习得研究方法 35 年: 回顾与思考[J]. 外国语 2004 (4).
- [7] 邢红兵. 基于联结主义理论的第二语言词汇习得研究框架[J]. 语言教学与研究 2009 (5).
- [8] 许尔茜. 留学生“V 着(zhe)”的习得研究[A]. 汉语作为第二语言习得的认知探索(邢红兵主编) [C]. 北京: 世界图书出版公司 2011.
- [9] 张银丹. 留学生形名搭配句法与语义关系习得的实验研究[A]. 汉语作为第二语言习得的认知探索(邢红

- 兵主编 [C]. 北京: 世界图书出版公司 2011.
- [10]周琳. 对外汉语教材同译词语及英语背景留学生使用偏误研究[D]. 北京语言大学硕士学位论文 2007.
- [11]Bates ,E. ,Elman ,J. ,Johnson ,M. ,Karmiloff-Smith ,A. ,Parisi ,D. ,Plunkett ,K. *Innateness and emergentism* [C]. In: Bechtel ,W. ,Graham ,G. (Eds.) *A Companion to Cognitive Science*. Blackwell ,Oxford ,1998.
- [12]Granger ,S. *From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora* [C]. In: Aijmer K. ,Altenberg B. and Johansson M. (eds) *Languages in Contrast. Text – based cross – linguistic studies*. Lund Studies in English 88. Lund: Lund University Press ,1996.
- [13]Granger ,S. *The computer learner corpus: A versatile new source of data for SLA research* [C]. In: Granger ,S. (Ed.) *Learner English on computer*. London: Longman ,1998.
- [14]Granger ,S. *A bird' s-eye view of learner corpus research* [C]. In: Granger et al. (eds.) *Computer Learner Corpora Second Language Acquisition and Foreign Language Teaching*. John Benjamins 2002.
- [15]Jiang ,N. Lexical Representation and Development in a Second Language [J]. *Applied Linguistics* 2000 (21) .
- [16]MacWhinney ,B. *The emergence of language* [M]. Mahwah ,NJ: Lawrence Erlbaum ,1999.
- [17]Nation ,J. S. P. *Teaching and Learning Vocabulary* [M]. New York: Heinle and Heinle ,1990.
- [18]Rumelhart ,D. & McClelland J. *On learning the past tenses of English verbs* [C]. In: James ,L. ,McClelland J. , David ,E. ,Rumelhart ,D. and the PDP Research Group (eds.) *Parallel Distributed Processing: Explorations in the Microstructures of Cognition: Psychological and Biological Models*. Cambridge ,MA: MIT Press ,1986.

A Corpus-based Approach to Second Language Lexical Acquisition

XING Hong-bing

(Beijing Language and Culture University ,Beijing 100083)

Abstract This paper analyzes the limitations of current studies ,such as in frequency effect ,error analyzes ,lexical awareness ,syntactic and semantic relations ,target language and native language relations. We should establish a corpus-based lexical knowledge database for the research of second language vocabulary learning. We recommend that the second language acquisition research should be organized around the principle of lexical knowledge is systematic ,developmental ,and research data should be analyzed using factor analysis method.

Key words: second language; lexical knowledge; lexical acquisition; corpus