

# 基于语料库的英语词汇教学: 理据与应用\*

王家义

(湖南工程学院 湘潭 411104)

**摘要:** 词汇是英语教学的重点和难点之一。近年来,随着计算语言学和语料库语言学的发展,语料库语言学的研究方法和技术手段被引入到英语词汇教学的研究和实践中,为英语词汇教学提供了新的途径和研究视角。本文从图式理论、词块理论和词汇语法理论的角度出发,为基于语料库的英语词汇教学寻求理论支持,探讨英语词汇教学在词频统计分析、词语搭配、语义韵律和词语辨析上的应用价值。

**关键词:** 语料库; 英语词汇; 理据; 应用

中图分类号: H319.3

文献标识码: A

文章编号: 1000-0100(2012)04-0127-4

## Corpus-based Approach to English Vocabulary Teaching: Motivation and Application

Wang Jia-yi

(Hunan Institute of Engineering, Xiangtan 411104, China)

The acquisition of vocabulary is one of the important and difficult areas of English teaching. In recent years, with the development of computational linguistics and corpus linguistics, research methods and technical means of corpus linguistics have been introduced to the study and practice of English vocabulary teaching, which provides a new approach and research perspective to it. This thesis analyzes the motivation of corpus-based English vocabulary teaching from the perspective of Schema Theory, Lexical Approach and Lexical Grammar, and explores its application values from the aspects of word frequency, collocation, semantic prosody and synonym discrimination.

**Key words:** corpus; English vocabulary; motivation; application

### 1 引言

王晶从多维度考察人的“语言能力”(王晶 2011)。就本文的主题而言,词汇量的大小直接决定着语言学习者的语言运用能力。虽然英语教师和学习者都非常清楚词汇在语言学习中的重要性,但长期以来英语课堂教学并没有给予词汇教学应有的地位。濮建忠认为,产生这种现象的原因如下:(1)对词汇知识关键要素的认识还不够充分;(2)缺乏有效地教授词汇的方法和手段(濮建忠 2003: 438)。语料库以其宏大的数据库为词汇教学提供可靠、客观的语料。语料库特有的语境共现界面可以提供大量检索项居中的实例,为激活归纳式学习策略提供前所未有的条件和支撑。本文探讨基于语料库的英语词汇教学理据,分析语料库在英语词汇教学中的实际应用价值。

### 2 基于语料库的英语词汇教学理据

#### 2.1 图式理论的理据

“图式是我们头脑中的关于普通事件、客体与情境的一般知识结构。”(王小明 1999: 62)作为认知基础,图式在大脑中一旦形成,会重新组织、理解和记忆以后获得的信息。人们在理解信息时,储存在大脑中的信息会将新信息与背景知识(已知概念、过去经历)联系起来。一旦输入信息与这些图式吻合,图式便起作用,完成对信息的接受、解码、重组和储存。图式理论强调人已有知识和知识结构对当前认知活动的决定作用,是“人们用已有的结构来记住新资料的一种方法”(桂诗春 2003: 5)。英语学习者头脑中已储存的知识对他们吸收新知识的方式和运用效果起关键作用。学习者可以通过建构图式的方式来

\* 本文系湖南省普通高校教改项目“基于语料库的大学英语词汇教学研究与实践”(湘教通[2010]243号331)的阶段性成果。

提高学习质量、培养熟练的外语能力。研究表明,语料库的语境共现可以为建构和重构图式的途径提供合适的语言材料。这样,用语料库设计出各类教学任务,引导学习者进行甄别、选择、分类、分析、综合,可以促使学习者不断提升和改善外语知识与能力,使早期习得的基本词汇及其搭配不固化。

## 2.2 词块理论的理据

自 1993 年 Michael Lewis 提出词块中心教学法理论 (Lexical Approach) 以来,词块在语言教学中的作用引起了研究者的广泛关注。Lewis 认为,语言是由语法化的词汇而不是词汇化的语法构成,倡导以输入大量真实语块作为语言学习的基本单位 (Lewis 1993: 89)。Wray 把词块定义为出现频率高、作为整体储存和使用的词语组 (Wray 2002: 9)。词块以整体形式存储在记忆中,使用时无需语法生成和分析,可直接整体提取使用。Nattinger 和 Delarrico 的研究表明,英语交际不仅仅通过单词或固定词组完成,90% 的自然交谈通过介于两者之间的半固定性语块实现 (Nattinger & Decarrico 2000)。研究表明,水平较高的学生能借用更多语块,语块运用能力和写作成绩相关。学习者能否掌握一个词的意义和用法,关键在于“能否掌握能够体现常用意义和用法的常用词块”(韩晓蕙 2011: 97)。词块教学的总体思路:将常用词放回到上下文中(而不是孤立出来),充分考虑使用时的关键语境,即提炼出能够充分体现典型类联接和搭配的词块加以注意和学习。(濮建忠 2003: 444) 基于语料库的英语词汇教学符合 Lewis 倡导的以输入大量真实语块作为语言学习的基本单位的理念。语料库可以实现搭配的自动提取,我们借助词性赋码的语料和合理的检索,能够挖掘很多句法结构的类联接特征。语料库将大量的实例以数据或语境共现的形式呈现在学习者面前,这有利于吸引学习者的注意力,强化学习者的记忆,帮助学习者利用语境获取语义和总结规律。

## 2.3 词汇语法的理据

在传统语法中,语法和词汇一直被视为语言基本模式的两个范畴,语法的地位远远超过词汇:前者是基本框架,后者仅是填充物,而语言的意义则来自于由语法先搭好的架构,然后再按语法规则填入符合一定词性的词。Sinclair 认为,这种“slot-and-filler”模式没能充分揭示语言的本质,这种过于简单的意义构建模式是“前计算机时代”不够充分和完善的语言观察手段所致 (Sinclair 2004: 165-169)。能够嵌入语法框架的词并非只要求词性相同,某种语法意义的实施也不是单单依靠句法形式。词汇和语法同等重要,而且要优先以词汇作为出发点来构建语言模式。在传统英语语法中,单个词一直被视为语言中最小的表意单位,但是,词汇语法理论却认为构建语言意义的最基本单位是词项。辛格莱提出词项多层意义

关系的描述框架。第一层关系是词汇搭配 (collocation),指词项与周边词的典型相伴形式,通常的相伴距离为词项前后 4 个词以内的范围。第二层关系是语法搭配或类联接 (colligation),指与词项相伴的语法结构,主要是语法词类的频繁组合。第三层关系是语义偏好 (semantic preference),指词项与周边词在语义上的频繁共选倾向,它关注惯常搭配词的语义特征或类别,如 sports 类, suffering 类等。第四层关系是语义韵 (semantic prosody),指由上述这些语义倾向所带出的语言使用者的立场、态度或感情取向特征,它散布于多个词的范围内,是形成整个词项意义的决定因素。Sinclair 以语料库为基础提出从语法形式和词汇相结合的方法来描述英语的观点。这种观点否定将语法和词汇分开描述的传统方式,为基于语料库的语言描述开辟了一条新路。同时, Sinclair 和 Renouf 还在大量语料库研究基础上提出词汇大纲的思路和设想,明确指出,英语教学的重点应该是语言中最常见的词形及其构成的核心用法模式、典型组合 (combinations)。(濮建忠 2003: 439) Sinclair 的词汇语法的理论的重要贡献在于,将短语从词汇学研究的边缘带回到核心。语料库语言学家在对语言的观察和描述方面与传统的语言学家不同。他们观察包括口语和书面语在内的使用中的真实语言,其描述单位是涉及结构和词汇的意义单位。这能够为英语教学提供准确和客观的依据。传统语法由于受到种种限制,对英语真实用法的描述在很多方面都不准确。英语教学以传统语法描述为依据同样不准确。

## 3 语料库在大学英语词汇教学中的应用

### 3.1 利用语料库进行词频统计分析

目前,英语词汇总量已经超过 200 万。(汪榕培 1999) 在英语学习过程当中,不可能且没必要全部掌握如此海量的词汇。英语学习者要真正学会和掌握的是人们自然运用语言中最有活力、能产性最强和最基本的词汇。基于语料库的词频统计分析正好适应这一需求。利用语料库进行词频统计分析,可以较精确地确定不同等级的高频词汇的范围与数量。据 Francis 和 Kucera 的统计,在一百多万词的 Brown 语料库中,频率居前 1000 的单词就可以覆盖普通文本的 72%; 频率居前 2000 的单词可以覆盖 79.7%; 频率居前 3000 的单词可以覆盖 84%; 频率居前 4000 的单词可以覆盖 86.8%; 频率居前 5000 的单词可以覆盖 88.7%; 频率居前 6000 的单词可以覆盖 89.9%; 频率居前 15851 的单词可以覆盖 97.8%。(Francis & Kucera 1982) 可见,英语最常用单词(词频在前 2000 的单词)大约占语篇的 80% 词汇,这表明教师在教学过程中应把有限的时间和精力投入到最常用的词汇上。若英语学习者能首先掌握前 2000 个左右的高频词汇,就可以基本了解一般文章的大意。因此,学习现实生活中真正高频

共现词项,可以减少词汇学习中出现的词语搭配的随意性与局限性,加速学习者对词语搭配知识信息较为全面的掌握与活化,提高词汇教学的质量和英语学习效率。

### 3.2 利用语料库进行词语搭配教学

搭配是“在文本中实现一定的非成语意义并以一定的语法形式因循组合使用的一个词语序列,构成该序列的词语相互预期,以大于偶然的几率共现”(卫乃兴 2001: 100)。词的意义不是孤立的,能从与它结伴同现的词中体现出来。词项的结伴规律、结伴词项间的相互期待与相互吸引、搭配成份的类联接关系等都是词语搭配的形式属性,都是词语搭配研究的重要内容。(卫乃兴 2002a: 101)

利用语料库进行词语搭配教学,学生可以直观地学习词语的典型搭配。如通过考察 *absolutely* 在 BNC 语料库中与形容词搭配的相关信息值,就可以发现与 *absolutely* 显著搭配的形容词分别为:表确定的词,如 *sure* ,*certain*;表正确的词,如 *right* ,*correct*;表美好的词,如 *brilliant* ,*wonderful* ,*delighted* ,*perfect* ,*fine* ,*gorgeous* ,*fantastic*;表必要的词,如 *necessary* ,*essential* ,*crucial* ,*vital* 等。

教师可以用语料库客观地分析学生的用词搭配。我们研究工科学生在英语写作中 *get* 的使用,结果发现:在“*get + adj*”结构中,*get* 作为“系词”使用(李洪儒 2006: 33)。据此,学生能用词具体准确,充分表达句子的意义,如 *get more beautiful* ,*get familiar with* ,*get addicted into* ,*get hooked on* 等。而对“*get + n*”结构,学生选择最多的搭配词是 *information* ,*news* ,*date* ,*mail* 等,但也有少量搭配词不地道或者属于中式英语,如 *get answer* ,*get feelings* ,*get the entertainment* 等。可见,凡是表达“得到、获取”之义时,学生都将 *get* 视为“万能动词”,认为该动词适合各种语境,能行使各种功能。这说明学生词汇量有限,不能借助不同词表达同一概念或意义;同时也反映出学生对 *get* 一词的内涵缺乏深入的了解。至于 *get* 的固定搭配,高频率出现的是 *get rid of* ,*get into the cyber trap* ,*get the best use of* ,*get away from* ,*get contact with* ,*getting on* 等。总的来说,学生对 *get* 一词的掌握不够深入和全面,表现为缺乏多样性和灵活性。(刘向红 王家义 2009: 122)

### 3.3 利用语料库进行语义韵教学

语义韵是语料库语言学的重要课题,可分为积极、中性和消极 3 类。在消极语义韵里,关键词吸引的词项几乎都具有强烈或鲜明的消极语义特点,使整个语境弥漫一种浓厚的消极语义氛围。积极语义韵的情况正好相反:关键词吸引的几乎都是具有积极语义特点的词项,由此形成一种积极的语义氛围。在中性语义韵里,关键词既吸引一些消极含义的词项又吸引一些积极含义或中性含义的词项,由此形成一种错综的语义氛围。因此,中性语义韵又可称做错综语义韵。绝大多数英语词的搭配

行为呈现出错综语义韵的特性,一些词项具有强烈的消极语义韵,另一些词项则有明显的积极语义韵。(卫乃兴 2002b: 300) 卫乃兴提出语义韵研究的一般方法:(1) 建立并参照类联接,用基于数据的方法研究;(2) 计算节点词的搭配词,用数据驱动法研究;(3) 用基于数据与数据驱动相结合的折中方法研究。学生通过观察节点词的索引行就能分析节点词的语义韵特征。以 *rather* 为例,其右搭配形容词/副词依序分别为 *superfluous* ,*fine* ,*dismal* ,*squalid* ,*ugly* ,*sad* ,*desolate* ,*exaggerated* ,*eerie* ,*mean* ,*indignant* ,*shamefaced* ,*pessimistic* ,*disappointing* ,*impatient* ,*too* ,*sordid* 等。可见,*rather* 修饰形容词、副词时,其搭配词绝大部分都是一些表示不好的、具有消极含义的词项。再如 *commit* 这个词在 BNC 语料库中右搭配词主要为 *suicide* ,*crime* ,*murder* ,*offence* ,*rape* ,*adultery* ,*arson* ,*incest* 等具有消极色彩的词,*commit* 的右搭配词具有明显消极语义韵特征。

### 3.4 利用语料库进行词语辨析教学

语料库语言学途径的同义词辨析主要通过考察同义词在不同语境中的词频分布差异,统计各搭配词与关键词共现的显著程度,观察检索行中呈现的同义词搭配特征,揭示其不同的类联结、搭配关系和语义韵等语言特征。得益于巨大库容的语料库提供的翔实证据,语料库索引可对同义词群提供丰富的用法和语境,使研究者能够比较和掌握同义词间细微的语义、语用差异,并对此进行客观、全面地描述。

#### 3.4.1 词频分布差异

词频分布差异主要考察同义词在不同语境中的使用频率差异。“同义词之间除了有定性差异外,在使用频率以及不同语境词频分布上都有较明显的差异,这种差异也可成为判断的依据之一,而且更适合作判断的前提,判断应优先考虑哪个词。”(潘璠 冯跃进 2000: 8) 以 *utterly* 和 *absolutely* 为例,这两个词在 BNC 语料库各文体中每一百万词次的宏观分布情况:(1) *absolutely* 在各语体中的使用频率均高于 *utterly*;(2) *utterly* 在小说中的使用频率最高,共出现 481 次,平均每百万词出现 30.23 次,*absolutely* 在口语中的使用频率最高,共出现 1889 次,平均每百万词出现 189.59 次。

#### 3.4.2 搭配差异

通过研究词的搭配差异可以区分同义词。具体讲,将关键词的所有搭配词从语料库中提取出来,采用统计手段测量各搭配词与关键词共现的显著程度,确定各词项之间在多大程度上相互吸引,概括出词项的典型搭配程度。

通过考察 *absolutely* 和 *utterly* 在 BNC 语料库中与形容词搭配的相互信息值,发现与 *absolutely* 显著搭配的形容词分别为:表确定的词,如 *sure* ,*certain*;表正确的词,如

right ,correct; 表美好的词,如 brilliant ,wonderful ,delighted , perfect ,fine ,gorgeous ,fantastic; 表必要的词,如 necessary , essential ,crucial ,vital 等。与 utterly 显著搭配的形容词主要为消极意义词,如 ridiculous ,miserable ,wrong ,disgraceful ,ruthless ,unexpected ,unable ,alien ,stupid ,inadequate; 表示安静、孤独的词,如 silent ,alone ,desolate; 表示相异的词,如 alien ,different 等。

### 3.43 语义韵差异

语义韵的提出为同义词辨析提供了新的维度。如通过考察 absolutely 和 utterly 的语义韵律,发现节点词 absolutely 的形容词搭配词具有明显的语义韵倾向,吸引的词汇是具有积极意义的搭配词汇(主要为表示确定、正确、美好、必要、重要的词),由此形成一种积极的语义氛围。在 Utterly 的前 20 个形容词主要搭配词中,有 12 个含有消极义韵(ridiculous ,impossible ,miserable ,wrong ,disgraceful ,dependent ,ruthless ,unexpected ,unable ,stupid ,desolate , inadequate) 9 个含有中性义韵(different ,dependent ,silent free ,new ,alien ,convincing ,alone ,convinced) 再把搭配词观察频数扩展至使用频率最高的前 50 个词,发现有 30 个搭配词为消极义韵词 6 个积极义韵词和 14 个中性义韵词。可见 utterly 的形容词搭配词整体上呈现消极语义韵特征:积极的占 12% ,消极的占 60% ,中性的占 28% 。

## 4 结束语

语料库应用于英语词汇教学可以帮助教师较科学地确定词汇教学的内容,找出教学重点,合理安排教学次序,使学生在有限的时间内学到最常用的语言知识,使教师及时、客观地纠正学生的用词错误,为制定教学大纲和编写教材提供参考。学生通过语料库调出大量语言实例后,根据其语境特征进行多层次分类和归纳,逐步挖掘出其中带有规律性的东西,从而激发他们的批判性思维和学习潜能,使词汇自主学习能力得到培养和提高。如今,语料库知识越来越普及,语料库的创建和使用越来越便利。互联网上不少语料库(如 BNC ,COCA 等)和一些语料库检索工具(如词频统计软件、自动词性附码软件和中文分词软件等)可以通过申请注册后免费使用,或下载后免费使用。我们相信,随着语料库、语料检索工具的不断创建、开发和更新,语料库的研究方法和技术手段将会不

断运用到英语词汇教学实践中,基于语料库的英语词汇教学研究也将得到不断的发展和深化。

### 参考文献

- 桂诗春. 记忆与英语学习[J]. 外语界, 2003(3).
- 韩晓蕙. 词块加工影响高校学生英语习得的研究[J]. 外语学刊, 2011(3).
- 李洪儒. 系词——人在语句中的存在家园——语言哲学系列探索之二[J]. 外语学刊, 2006(2).
- 刘向红 王家义. 基于语料库的工科学生英语写作用词特点透视[J]. 外语学刊, 2009(2).
- 潘 璠 冯跃进. 基于语料库的同义词差异性特征调查[J]. 山东外语教学, 2000(4).
- 濮建忠. 英语词汇教学中的类联接、搭配及词块[J]. 外语教学与研究, 2003(6).
- 汪榕培. 英语词汇探索[M]. 上海: 上海外语教育出版社, 1999.
- 王 晶. 解析言语能力发展的建构过程[J]. 外语学刊, 2011(3).
- 王小明. 图式理论与句子教学[J]. 华东师范大学学报(教育科学版), 1999(2).
- 卫乃兴. 词语搭配的界定与研究体系[M]. 上海: 上海交通大学出版社, 2001.
- 卫乃兴. 基于语料库和语料库驱动的词词语搭配研究[J]. 当代语言学, 2002a(2).
- 卫乃兴. 语义韵研究的一般方法[J]. 外语教学与研究, 2002b(4).
- Francis, W. N. & H. Kucera. *Frequency Analysis of English Usage* [M]. Boston: Houghton Mifflin Company, 1982.
- Lewis, M. *The Lexical Approach* [M]. London: Language Teaching Publications, 1993.
- Nattinger, J. & J. DeCarrico. *Lexical Phrases and Language Teaching* [M]. Shanghai: Shanghai Foreign Language Education Press, 2000.
- Sinclair, J. *Trust the Text: Language, Corpus and Discourse* [M]. London: Routledge, 2004.
- Wray, A. Formulaic Sequences in Second Language Teaching: Principles and Practice [J]. *Applied Linguistics*, 2002(21).

收稿日期: 2011-10-13

【责任编辑 王松鹤】