语料库与英语语言特征研究

Corpora and English Language Features

吴军 晁宏晏 苏莹 著

新华出版社

语料库与英语语言特征研究

吴　军　晁宏晏　苏　莹　著

新　华　出　版　社

图书在版编目(CIP)数据

语料库与英语语言特征研究/吴军,晁宏晏,苏莹著. —北京:新华出版社,2012.12

ISBN 978-7-5166-0200-3

Ⅰ.①语… Ⅱ. ①吴… ②晁… ③苏… Ⅲ.①英语—语言学—研究 Ⅳ.①H31

中国版本图书馆 CIP 数据核字(2012)第 287432 号

语料库与英语语言特征研究

作　　者:吴　军　晁宏晏　苏　莹

出 版 人:张百新　　责任印制:廖成华

责任编辑:徐　光　　装帧设计:张　斌

出版发行:新华出版社

地　　址:北京石景山区京原路 8 号　　邮　　编:100040

网　　址:http://press.xinhuanet.com　　http://www.xinhuapub.com

经　　销:新华书店

购书热线:010—63077122　　中国新闻书店购书热线:010-63072012

照　　排:安阳师范学院印刷厂照排部

印　　刷:安阳师范学院印刷厂

成品尺寸:170mm×235mm　1/16

印　　张:18.5　　字　　数:332 千

版　　次:2013 年 3 月第一版　　印　　次:2013 年 3 月第一次印刷

书　　号:ISBN 978-7-5166-0200-3

定　　价:39.00 元

图书如有印装质量问题,请与出版社联系调换:(010)63077101

前　言

近年来，语料库语言学已逐渐成为语言研究的主流。这种变化不仅反映了语言学研究领域中思想观念的更新，而且还反映了现代语言研究向着更具描述性、更趋科学性的方向发展的趋势。

语料库问世之前，语言特征一直由本族人的直觉来提供，可是，针对语言使用频率和语域间的语言差异等语言特征，本族人很难有精确的感知。只有大规模、有代表性的语料库才是唯一可靠的语言信息来源。语料库具有前所未有的巨量语言信息储备、高速精确的计算机提取方式和鲜明突出的语境共现界面，因此对语言教学及研究具有革命性影响。本书建立在五个大型语料库的基础上，对英语语言特征进行了详尽研究，这五个大型语料库是：COBUILD 语料库、朗文语料库、英国国家语料库、美国当代英语语料库和历时美语语料库。本书的研究覆盖以下几个方面：

（1）频率和语域分布的统计。本书基于语料库对英语的使用频率和语域分布提供了全面而可靠的统计。对于教师、学生、材料撰写者以及学术研究者来说，了解语言中常见以及罕见的词语和结构是十分有益的，频率信息可以有效指导人们的语言学习、教学及研究；语言学家 Biber 指出，当人们在不同语域中使用一种语言时，他们是用语言做非常不同的事情。每个语域中的语言结构都有其不同的特点，对语域不加区分的整体语言描述常常遮掩了不同语域中语言特征的重要差别。了解不同语域中语言的使用情况有助于人们提高不同场合中的交际能力。

（2）结构和词汇的联系。长期以来，语法结构研究一般事实，词汇处理具体事实，结构和词汇之间缺乏联系，导致许多抽象规则无法具体运用。基于语料库的语言研究架起了语法和词汇结合的桥梁。本书基于语料库将语言结构和词汇结合起来，提供了与结构相关的词汇信息，实现了横向组合和纵向聚合的联系；

（3）词串研究。词串是在大规模语料库的研究基础上突显出的新的语言现象。它不属于任何传统的语言结构，然而词串重复出现的频率体现了其重要地位。语言学家 Biber 认为，要造出自然、地道的英语不只是要用结构正确的句

子,还要用经常使用的符合习惯的表达语。在这方面,词串为学习者提供了重要信息。本书研究了词串的概念、功能及使用情况。

(4)真实例句。语料库语言学家 Sinclair 指出,重要的是让读者只学习、研究真正的语言实例,特别是把例句作为使用楷模时,更是如此。用 Birds sing 之类的例句来说明简单的主谓句,从语法理论来看,这样的句子是正确的,但如果用语料库检索一下这样的句子,就不难发现它们是罕见的。显然,杜撰的句子很难提高学习者的交际能力。本书中的例句全部来自语料库,真实可靠,可用于语言教学和研究。

(5)语料库的应用。利用语料库辅助语言教学的先行者是 Tim Johns 和 Chris Tribble。他们开创的"数据驱动学习(DDL)"从理论上看是一种很好的英语学习方法,但是在实际使用过程中存在着诸多问题,因此至今未形成太大气候。如何在教学中高效利用语料库资源是值得研究的课题。本书探讨了语料库直接与间接应用相结合的综合应用模式。

本书由吴军、晁宏晏、苏莹共同撰写完成:其中吴军负责全书的统筹和整理工作,并撰写了第一章、第二章中的 2.1 节、2.2 节、2.3 节、2.4 节、2.5 节、2.6 节、2.7 节、2.8 节、第三章、第四章,共计 14.8 万字;晁宏晏撰写了第二章中的 2.9 节、2.10 节、2.11 节、2.12 节、2.13 节、2.15 节,共计 9.6 万字;苏莹撰写了第二章中的 2.14 节、2.16 节、2.17节、2.18 节、2.19 节、2.20 节、2.21 节、2.22 节、2.23 节、2.24 节、2.25 节、2.26节、2.27 节,共计 8.6 万字。

本书如有未尽之处,欢迎读者批评指正。

吴　军　晁宏晏　苏　莹
安阳师范学院外国语学院
2012 年 11 月

试读结束,需要全本PDF请购买 www.ertongbook.com

目 录

第一章 语料库语言学简介 …… (1)
1.1 基本概念 …… (1)
1.2 发展简史 …… (1)
1.3 语料库主要类型 …… (3)
1.4 基于语料库的语言特征研究意义 …… (5)
第二章 基于语料库的英语语言特征 …… (10)
2.1 名词 …… (10)
2.2 代词 …… (18)
2.3 限定词 …… (26)
2.4 形容词 …… (39)
2.5 名词属格 …… (56)
2.6 名词前置修饰语 …… (67)
2.7 数词 …… (73)
2.8 名词后置修饰语 …… (81)
2.9 动词 …… (90)
2.10 补足语 …… (107)
2.11 “动词1+动词2”结构 …… (117)
2.12 句子种类 …… (124)
2.13 否定结构 …… (139)
2.14 情态助动词 …… (151)
2.15 动词时态 …… (167)
2.16 状语 …… (190)
2.17 副词 …… (196)
2.18 间接引语 …… (203)
2.19 名词性分句 …… (210)
2.20 关系分句 …… (213)
2.21 状语分句 …… (219)

2.22 非限定分句 …… (225)
2.23 并列结构 …… (230)
2.24 被动结构 …… (232)
2.25 存在句 …… (237)
2.26 IT – 句型 …… (239)
2.27 倒装结构 …… (241)
第三章 基于语料库研究的特殊结构: 词串 …… (246)
3.1 词串的概念及界定 …… (246)
3.2 词串的功能及使用情况 …… (247)
3.3 会话中的词串类型及实例 …… (250)
3.4 学术文章中的词串类型及实例 …… (253)
第四章 语料库在英语教学中的综合应用 …… (257)
4.1 历史背景 …… (257)
4.2 综合应用 …… (263)
4.3 语料库语言学发展方向 …… (283)
参考书目 …… (289)

第一章 语料库语言学简介

1.1 基本概念

语料库(corpus,复数为 corpora) 是指为研究语言,用计算机处理和储存的书面和口头的语言材料。语料库一词来源于拉丁语,本意为 body。今天语料库指的是一个“电子文本集”。一个小型文本集并不是真正意义上的语料库。真正的语料库是一个按照一定的采样标准采集而来的、能够代表一种语言或者某语言的一种变体或文类的电子文本集。可以说,一个语料库由若干个电子文本构成,而这些电子文本作为一个整体可以代表某语言或某语言的某种变体或文类。因此,以一个语料库为数据源进行的研究可以看作是对该语料库所代表语言、语言变体或文类的研究,研究所得到的结论可以推广到整个语言、语言变体或文类。

一些人认为语料库语言学是一个独立的学科,它有自己独到的理论体系和操作方法。由于语料库语言学立足于大量真实的语言数据,对语料库做系统而穷尽的观察和概括所得到的结论对语言理论建设具有无可比拟的创新意义。而另外一些研究者认为语料库语言学并非语言学的又一个分支学科,而是一种研究方法,这种方法基于大量的真实语言,可以用来回答通过其它途径很难回答的问题,从而极大地丰富已有的研究方法。语料库语言学以大量精心采集而来的真实文本为研究素材,主要通过概率统计的方法得出结论,因此语料库语言学从本质上讲是实证性的。

1.2 发展简史

最早的语料库可以追溯到18 世纪,然而直到20 世纪50 年代后期在计算机技术的推动下语料库才逐渐发展起来。如果从语料库规模、语料库收集的特点以及动机等因素考虑,可以将国外英语语料库语言学的发展历史归纳为以下四个重要阶段:

(1) 从20世纪60年代起的小型语料库,其规模通常是100万词次或者更少,如BROWN和LOB语料库。后来有了这些语料库的词性附码版本,如BROWNTAGGED和LOBTAGGED,以及语音标注版本,如LLC,以上三者可称为早期的三大经典语料库。

(2) 从20世纪80年代起的大型语料库,其规模是以往的数十万乃至数百万倍,如730万词次的Cobuild语料库很快发展成1.67亿词次的BoE语料库。

(3) 从20世纪90年代末起的动态型语料库,其特点之一是对早期语料库实行后期的内容更新,如20世纪60年代的BROWN和LOB语料库更新为20世纪90年代的FROWN和FLOB;特点之二是建立开放性的、滚动式发展的历时性语料库,如自1998年起延续15年一直在扩展的英国Independent和Guardian报刊语料库。

(4) 从2005年起的电子网络语料库,其特点是在国际互联网上设置检索引擎,将互联网上的语言信息作为一个巨大的、动态的和开放的语料库,如WEBCORP。

语料库语言学发展的动因有三:一是科学研究的动机,即由好奇心引发的科学论证精神;二是语言使用的需要,如出版机构、语言教学的需求;三是人类特有的创新的本能。总结起来,语料库语言学的发展反映了人类对知识的渴望,对语言使用的需求和现代科学技术发展的推动力。

在国内,语料库语言学起步于20世纪80年代,如上海交通大学建立的国内首个百万词次的科技英语语料库JDEST(Yang. 1986)。进入21世纪以来,语料库语言学在国内逐步推广起来,近年来发展尤为迅猛,呈现出以下特点:

(1) 注重建设外语学习者的中介语语料库。先后建成并有广泛影响的有《中国学习者英语语料库(CLEC)》(桂诗春和杨惠中,2003)以及《中国学生英语口笔语语料库(SWCCL)》(文秋芳等,2005/2009)等。

(2) 注重建设汉语语料库以及汉语与外语匹配的双语或平行语料库。例如:《国家现代汉语语料库》(国家语委,2009)以及《英汉双语语料库》(王克非,2003)等。

(3) 注重建设外语教学语料库。例如:华南师范大学外国语言文化学院在2000年就研制出版了《中学英语教育语料库》(华南师范大学外文学院,2000)。近年来还出现了基于某个语域或某个专业学科的英语教学而建设的各类教育或学术语料库,如《基础英语教材语料库》(和安平和郑旺金,2009);《商务英语语料库》、《中医英语语料库》等也在建设中。

语料库未来的发展方向有可能是由后互联网时代的网络技术支持的更为即时的同步的、多模态的以及全球整合型的巨量语料资源库。

1.3 语料库主要类型

因研究目的的不同,语料库也有多种类型,以代表各种各样的语言、语言变体或文类。常见的语料库类型主要有:

1.3.1 通用语料库(general corpus) :

广泛采集某语言的口、笔语形式,取样是尽可能考虑口、笔语的主要社会变体、地域变体、行业变体等各种变异及语言使用的各种场合之间的平衡,力求最好的代表一种语言的全貌而建成的语料库。通用语料库一般较大,常常达到数亿词次,代表性的英语通用语料库有英语国家语料库(British National Corpus, BNC) 、英语文库(Bank of English, BoE) 、美国国家语料库(American National Corpus, ANC) 等。通用语料库是描述语言全貌、编制工具书、核查语言用法等最理想的语料。此外,通用语料库还常常被用作参考语料库,以方便我们发现某些专门语料库的语言特点。

1.3.2 专用语料库(specialized corpus) :

专用语料库又称专题语料库(special purpose corpus) 。与通用语料库相反,出于某种特定的研究目的,人们常常只收集某特定领域的语料库样本建成语料库,此类语料库称为专用语料库。在实际研究中,人们常常将专用语料库与通用语料库进行对比,来分析特定领域内语言的特点。此外,专用语料库也可作为编制专门领域工具书的理想语料。

1.3.3 共时语料库(synchronic corpus) :

由同一时代(主要是当代) 的语言使用样本构成的语料库称为共时语料库。共时语料库是相对历时语料库而言的。基于不同时代的语言所建成的多个共时语料库可以构成一个历时语料库。

1.3.4 历时语料库(diachronic corpus) :

收集不同时代的语言使用样本构建而成的语料库称为历时语料库。历时语料库是观察和研究语言变化时常用的语料库。对历时语料库进行分解可以得到多个共时语料库。赫尔辛基英语文本语料库(Helsinki Corpus of English Texts) 是一个典型的英语历时语料库。

1.3.5 口语语料库(spoken corpus) :

口语语料库常常包括由口语转写而来的文本,有时也包括语音文件。因为取样和转写的困难,口语语料库的文本容量很难达到笔语语料库的规模。将口语语料库和通用语料库进行对比,可以有效地发现口语特征。为了方便口语研究,人们常常对口语语料库中的语音、语调、停顿、重复、修正等口语特征进行标注。

1.3.6 笔语语料库(written corpus) :

笔语语料库取材于书面语,常常包括书籍、报刊、书信、学术论文等常见笔语形式。由于笔语文本较容易收集,笔语语料库的容量一般较口语语料库的容量更大。

1.3.7 本族语者语料库(native speakers' corpus) :

本族语者语料库中所收集的语言使用样本,全部源自于本族语者。本族语者语料库区别于非本族语者语料库和学习者语料库,在分析非本族语者或学习者语言使用特点时,经常以本族语者语料库作为参照。

1.3.8 学习者语料库(learner corpus) :

学习者语料库是由非本族语学习者语言使用样本构成的语料库。学习者语料库又可分为口语语料库和笔语语料库。国际上影响较大的学习者语料库有比利时学者 Sylviane Granger 等人于 20 世纪 90 年代初建立的英语学习者国际语料库(International Corpus of Learner English, ICLE) 和鲁汶英语中介语国际数据库(Louvain International Database of Spoken English Interlanguage, LINDSEI) 等。国内较有影响的学习者语料库有中国学习者英语语料库(Chinese Learners' English Corpus, CLEC) (桂诗春、杨惠中 2003) 、中国学生口笔语语料库(Spoken and Written Corpus of Chinese Learners, SWECCL 1.0 & SWECCL 2.0) (文秋芳等 2005; 文秋芳等 2008) 、中国学习者英语口语语料库(College Learners' Spoken English Corpus, COLSEC) (卫乃兴等 2005) 、中国大学生英汉汉英口笔译语料库(Parallel Corpus of Chinese EFL Learners, PACCEL) (文秋芳、王金铨 2008) 、CEM(Corpus for English Majors) 语料库(中国高校英语专业多语种语料库建设和研究项目组 2008) 等。

试读结束, 需要全本PDF请购买 www.ertongbook.com

1.3.9 单语语料库(monolingual corpus) :

单语语料库中的语料库来自于同一种语言,如英语语料库、汉语语料库等。

1.3.10 平行/双语语料库(parallel/bilingual corpus) 和多语语料库(multilingual corpus)

平行/双语语料库中的语料来自于两种语言,而且相互对应,即一种语言是另一种语言的译文。双语语料库建设中的一个重要环节是两种语言间的对齐(alignment) 问题。目前,大多数双语语料库都进行了句之间的对齐,也有人尝试词语间的对齐和意义单位之间的对齐。双语语料库对翻译研究和机器翻译研究具有重要价值。北京外国语大学王克非教授主持建立的英汉双语平行语料库是国内较有影响的英汉汉英平行语料库。

多语语料库中的语言使用样本取自于多种语言。如 Europarl Parallel Corpus(European Parliament Proceedings Parallel Corpus) 收集了欧洲议会的多种语言文集,将 11 种语言进行对齐处理,该语料库可以从网上免费下载。

1.4 基于语料库的语言特征研究意义

自语料库产生以来,它在语言领域的研究成果首先在词典编纂方面取得了突破,基于语料库的英语词典现已成为国外出版的英语学习词典的主流,如根据 Cobuild 语料库编纂的词典《柯林斯合作英语词典》(*Collins CBUILD English Language Dictionary*) ,以朗文语料库(Longman Corpus Network) 为基础编纂的词典《朗文当代英语词典》(*Longman Dictionary of Contemporary English*) 。基于朗文语料库的《朗文当代英语词典》把语域分布和频率信息应用在以下几方面:

(1) 标出了英语口语和笔语中最常见的 3000 词。例如,词条 bother(v) 旁边标有 S1 W3,表示该词属于口语中最常用的一千词,笔语中最常用的三千词,即该词在口语中比在笔语中常用。由外研社出版的该词典的英汉双解本(在我国普遍使用) 使该标示更形象化,S 被一个嘴形简笔画代替,W 则被一个铅笔形简笔画代替。

(2) 在一些重要词的后面附上图表,作出使用率比较。例如: 在 enter 词条下用图表显示 enter 和 go/come in 在英语口语和笔语中的使用频率,该图表表明 go/come in 在口语中的使用频率是 enter 的 20 多倍,而在笔语中 enter 比 go/come in 更常用。有些图表列出某个词在各种句型、搭配中的使用频率,例如 need 词条下,分别标出 need sth,need to do sth,need sb/sth to do sth…句型的使

用频率。

(3) 按使用频率排列词条的义项、成语等顺序。例如,在 lookout 词条下,排在最前面的是 be on the lookout for 这个成语,接着是 keep a lookout,第三才是单词本身的意义。这种排序表明,lookout 最常见的是以成语形式出现,因此掌握成语比掌握单词本身更重要。

其次,语料库在语言领域的研究成果还体现在基于语料库的语法著作,例如 John Sinclair 根据 Cobuild 语料库(The Bank of English) 编写的语法书《英语语法大全》(*Collins CBUILD English Grammar*);Douglas Biber 等语言学家编写的颇具影响力的语法著作《朗文英语口语和笔语语法》(*Longman Grammar of Spoken and Written English*)。

具体而言,研究基于语料库的语言特征具有以下几点意义:

1.4.1 获取真实、鲜活的语言

Sinclair 指出,"重要的是让读者只学习、研究真正的语言实例,特别是把例句作为使用楷模时,更是如此。"语料库研究者可以从语料库中获取真实的、英语本族人正在使用的语言。这一点正是传统的语言研究者所欠缺的。用杜撰的句子说明语言现象、做语言分析会出现与真实实际使用相脱节的现象。如用 Birds sing 之类的例句来说明简单的主谓句,从语法理论来看,这样的句子是正确的,但在实际生活中会不会有人用这样的句子却全然不知。如果用语料库检索一下这样的句子,就不难发现它们是罕见的。

真实的语言有助于英语研究者总结某一语言结构的实际使用情况。根据章振邦主编的《新编英语语法教程》,drunk 用做表语和 - ed 分词,drunken 用做定语。吴丽英、王凤元(2004) 用 Cobuild 语料库检索 drunk 和 drunken 做表语和定语的搭配用法后得出的结论是: drunk 既可以做表语,也可以做定语,但是做定语时出现的频率要远远小于做表语时的频率,而且做定语时所引起的内涵意义有别于 drunken; 相比之下,drunken 的用法要局限得多,只能用做定语。可见没有语料库的帮助,语言研究者是很难把握 drunk 和 drunken 的真实用法以及它们之间的细微差别。

宋京生(2006) 用英国国家语料库(BNC) 检索了不同类型的存在句的分布频率,发现存在句在 BNC 中的分布频率虽然很高,但除了表示否定意义的"There + be + no"结构外,主要是以肯定句的形式出现。在检索出的 186030 例存在句中,否定句大约只占总数的 1.1%,这确实是我们始料未及的。在我国的语法教材中,一般都会出现否定形式的存在句,有些甚至还设计了把存在句由肯定句变为否定句的练习。对照中国学习者英语语料库(CLEC) 中存在句的

分布频率,发现中国学生有超用否定式存在句的倾向。由此可见,脱离语料库的语言研究和教学有可能是不可靠的。

从语料库中检索到的一组真实例句还可以共同说明在单个句子里不明显的语言点,例如,下面是从 Cobuild 语料库中检索到的一组含有短语动词 break out 的例句:

The moment work stops, disorder is liable to break out.

If he gets promoted, all hell will break out.

This caused an epidemic to break out among them.

This final destructive fever had to break out somewhere.

从这一组句中可以看出只有坏事才用 break out。这些从真实例句中总结出的语言韵很难从杜撰例句中概括出来,即使概括出要点也是不可靠的。

1.4.2 使语法和词汇得以联系

语言学家对语法和词汇各自的领域做过这样的规定: 语法研究一般事实(general facts) ,词汇处理具体事实(specific facts) 。长期以来,语法和词汇之间缺乏联系。许多规则似乎抽象得无法具体运用,因此对学习者掌握语言的帮助有限。这是语法教学不能取得理想效果的主要原因之一。基于语料库的语言研究架起了语法和词汇结合的桥梁。根据对语料库的研究和分析,Sinclair, Biber 等语言学家在描述语法时加入了大量的与某一语法形式相关的词汇信息,开辟了一条英语语法描述的新路。例如,It 的一种用法是做主语表示天气,《新编英语语法教程》(第四版) 提供了两个例句:

It is very warm and wet in South China these days.

It was very cold; it snowed and grew dark.

以上两句用到了两种句子结构: "It + v"和"It + be + C",补足语 C 可能是单独的形容词,也可能是形容词加表示一段时间的名词。可是,哪些动词和形容词常用于这两种结构却不得而知,学习者学了结构却不能灵活运用。与之不同的是,Sinclair 根据 Cobuild 语料库提供了与结构相联系的词汇信息。语言学家 Biber 根据朗文语料库的检索结果列出了经常用于现在时和过去时的动词。

这些与结构相关的词汇信息在传统的语法书上是看不到的。结构与词汇相联系的语法描述无疑会给语言研究和教学提供非常必要的信息,使学习者不仅熟悉了语法规则而且了解了语法规则的应用范围,避免学习了规则却一用就错的现象。语法和词汇的结合实际上是将横向组合(Syntagmatic relations) 和纵向聚合(Paradigmatic relations) 联系了起来,学习者可以根据交际的需要为特定

的句法结构选择适当的词，或者选择适当的词放入适当的结构，以表达思想、相互交流，逐渐达到自由表达的学习目标。

1.4.3 对语言现象的分布和频率提供全面而可靠的统计

对于教师、学生、材料撰写者以及学术研究者来说，了解语言中常见以及罕见的词语和结构是十分有益的。长期以来，这样的信息一直由本族人的直觉来提供，可是，针对语言使用频率的差异，本族人很难有精确的感知，只有大规模、有代表性的语料库才是唯一可靠的频率信息来源。其次，基于语料库的语言研究让我们了解了语域之间语言特征的差异。在这方面，本族人的直觉更不可靠了。朗文语料库的检索结果显示，很多对整体英语的综合描述不完全，甚至使人误解或不准确。每个语域中的语言结构都有其不同的特点。例如：名词和介词短语在新闻中的使用率大大超过会话中的使用率，而动词和副词在会话中要常见的多。再比如：人们通常认为会话中的语法结构简单，然而，检索结果显示，会话中的讲话人使用一些相对复杂的语法结构，如复杂的关系从句。Biber 等语言学家指出，当人们在不同语域中使用一种语言时，他们是用语言做非常不同的事情。对语域不加区分的整体语言描述常常遮掩了不同语域中语言特征的重要差别，实际上，这样的综合描述不能体现任何语域中的语言特征。因此，有必要对语料库中的每个语域进行描述、分析，以便获得各个语域中语言的真实使用情况。

例如，传统的语法教材中都会有这样一条规则：当 that - 子句是直接宾语或补语时，连词 that 在非正式用法里经常省略，形成不带 that 的子句。可是，“经常省略”经常到什么程度，正式语体中的省略程度又如何，学习者不知道。基于语料库的语言研究为学习者提供了全面而可靠的分布和频率信息统计。了解了这些分布和频率的检索结果后，学习者对 that - 子句省略的用法无疑会有更深入、更准确、更全面的把握。

1.4.4 突显出新的语言现象

词、短语、分句是语法书中涉及的结构层次。然而，通过对语料库的观察和研究，语言学家发现一些词经常成串出现，例如：

…an increase in the…　　…the extent to which…

…the fact that the…　　can I have a…

I don't know how…　　I thought that was…

这些串在一起的词既不是短语或固定搭配，也不是分句，实际上它们不属于传统语法书中描述的结构。语言学家 Biber 将这种非驴非马的经常出现的语

言现象称为“词串”(lexical bundle) ,词串的重要之处在于其重复出现的频率。在朗文语料库的会话和学术文章两个语域中,大量的语段由这些词串构成: 仅三词词串在会话中的每百万词中就出现 8 万多次,在学术文章中的每百万词中出现 6 万多次; 会话中,大约 30% 的词出现在词串里,学术文章中,大约 21% 的词出现在词串里。这样高度复现的语言片段,不能不引起语言学家的注意。由于词串结构的特殊性,传统的语法书都不曾提及或讨论词串的特征,正是基于大规模语料库的研究才把它的存在突显出来。这种语言现象是否属于语法研究的范围? Biber 等语言学家认为它属于语法研究的范围,因为“语法不仅是对抽象类别和结构的研究,也是对这些类别和功能中的具体的词以及它们具体的功能的研究”。语料库的研究使我们注意到了某些一直被忽视的真实语言中的现象。

由于语料库具有前所未有的巨量语言信息储备、高速精确的计算机提取方式和鲜明突出的语境共现界面,因此对语言教育发展具有革命性影响。研究语料库中的语言特征,其现实意义是显而易见的。本书建立在五个大型语料库的基础上,对英语语言特征进行了详尽研究。这五个大型语料库是 COBUILD 语料库、朗文语料库、英国国家语料库(BNC) 、美国当代英语语料库(COCA) 和历时美语语料库(COHA) 。本书的第二章基于以上语料库,对英语语言现象的语域分布和使用频率进行了统计,将语言结构和词汇联系起来,列出了某一结构的常用词语; 此外,本章节中的例句全部来自语料库,真实可靠。第三章基于语料库研究了词串这一新的语言现象。第四章探讨了语料库在英语教学中的综合应用模式。

第二章　基于语料库的英语语言特征

2.1 名词

名词用来指明人和物。

2.1.1 可数名词

可数名词是指可以计数的名词;可数名词有单、复数之分,且能与数字和某些限定词连用,如: a(n) , many, few, these, those, several, etc (如: a cow, two cows, many cows, several cows) 。

2.1.2 不可数名词

不可数名词是指无法用数目计算的名词,且词形不随数目的多少而变化。最典型的不可数名词都是单数形式,但也有一些复数名词同样不随数目的多少而变化,也不和数字连用。

2.1.3 集体名词

集体名词表示若干个个体组成的集合体。一些集体名词是可数名词,一些是不可数名词。可数的集体名词用法如个体名词。不可数的集体名词没有复数形式。

(1) 像 people, police, staff, cattle, militia, poultry, vermin 之类的集体名词通常表示复数,做主语时谓语动词要用复数:

People don't belch in this house.

Police are appealing for help from anyone who witnessed the incident.

When staff are absent, a class is split between other teachers.

People 通常和表示复数的量词连用 (many people, ten people) ,但如果指集体中的某一个体,就应使用相应的个体名词: a man/woman/person。然而,当 people 表示"民族,部落,种族"时,就可当做一般的可数名词来用:

试读结束，需要全本PDF请购买 www.ertongbook.com