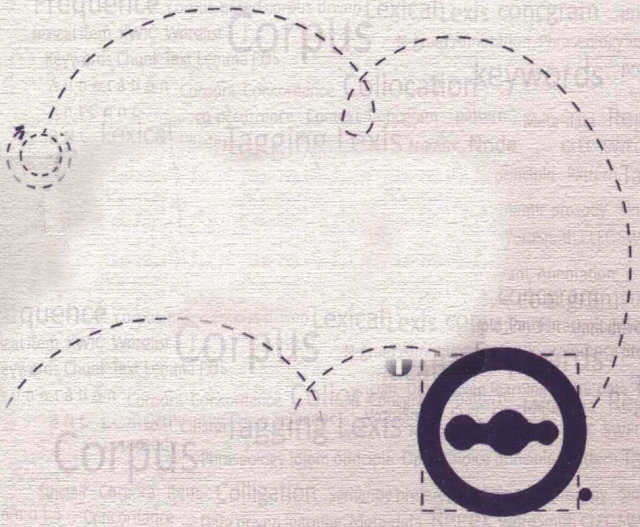
外研社基础外语教学与研究丛书·英语教师发展系列

主　编　程晓堂
副主编　韩　刚　林　立

语料库辅助英语教学入门

何安平 著



外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

外研社基础外语教学与研究丛书·英语教师发展系列
主　编　程晓堂
副主编　韩　刚　林　立

语料库辅助英语教学入门

何安平 著

外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS
北京 BEIJING

图书在版编目(CIP)数据

语料库辅助英语教学入门 / 何安平著. — 北京：外语教学与研究出版社，2010.12
(外研社基础外语教学与研究丛书. 英语教师发展系列)
ISBN 978-7-5135-0387-7

Ⅰ. ①语… Ⅱ. ①何… Ⅲ. ①英语—教学研究 Ⅳ. ①H319.3

中国版本图书馆 CIP 数据核字 (2010) 第 243093 号

出 版 人：于春迟
项目负责：范海祥
责任编辑：范海祥
执行编辑：郑加丽
封面设计：袁 璐
出版发行：外语教学与研究出版社
社　　址：北京市西三环北路 19 号 (100089)
网　　址：http://www.fltrp.com
印　　刷：三河市北燕印装有限公司
开　　本：650×980 1/16
印　　张：11.5
版　　次：2010 年 12 月第 1 版 2010 年 12 月第 1 次印刷
书　　号：ISBN 978-7-5135-0387-7
定　　价：25.00 元 (附赠 CD-ROM 一张)

* * *

基础教育出版分社：
地　　址：北京市西三环北路 19 号 外研社大厦 基础教育出版分社 (100089)
咨询电话：010-88819666 (编辑部)/88819688 (市场部)
传　　真：010-88819422 (编辑部)/88819423 (市场部)
网　　址：http://www.nse.cn
电子信箱：beed@fltrp.com 或登录 http://www.nse.cn (留言反馈)栏目
购书电话：010-88819928/9929/9930 (邮购部)
购书传真：010-88819428 (邮购部)

* * *

购书咨询：(010)88819929　　电子邮箱：club@fltrp.com
如有印刷、装订质量问题，请与出版社联系
联系电话：(010)61207896　　电子邮箱：zhijian@fltrp.com

物料号：203870001

外研社基础外语教学与研究丛书

——英语教师发展系列

专家委员会

前　言

本书适合作为语料库辅助英语教学的入门课程教材，对象是大、中、小学的英语教师以及高等院校相关专业的本科生和研究生。

现代信息技术对教育发展具有革命性影响，语料库是其中的重要组成部分。它以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取胜，一方面为语言教育教学提供大量优质资源，另一方面可创设大信息量、多维演示的立体教学和人机互动的优质教学平台。我国目前高等教育和基础教育领域的计算机普及率增长很快，但在利用语料库这个巨大而丰富的语言文化资源来促进外语教育教学方面还相对滞后；改变这一现状的重要途径是依靠广大语言教师来推广和普及语料库在外语教学中的使用。作为一名高师院校教师，笔者更感到责无旁贷，故将十几年来在华南师范大学利用语料库辅助英语教学的理论探索和实践经验汇集成书，目的是把语料库变成广大语言教师和高校学生进行语言学习、研究和教学的必备资源和工具。本书力求凸显以下理念：

（1）突出语言理论与语言事实相结合的科学精神，将基本理念和基础技术的教学落实到具体的案例演示和解读中。

（2）突出“精讲多练，在实践中探究”的现代学习理念。编写体例包括文献研读指引、实例演示述评、工作坊、教学反思和拓展性学习等版块，努力体现“以学生为主体，以学习为中心，注重实践，注重创新能力培养”的现代教育思想。

（3）实践纸质教材、电子光盘和互联网相结合的立体化现代教学资源理念。书中不仅有理论介绍、实例图示和分析，还提供批量网址，指

引学习者通过互联网获取更多的语料资源。本书附带光盘，内含若干免费语料库检索工具和一批微型语料文本，供书中40个工作坊上机使用。学习者可以按照工作坊的练习指令亲身体验语料库检索和分析，并且学会建设个人小型语料库来辅助教学。

本书除第二章和第六章分别由杨素香（河南理工大学）和杨文坤（华南师范大学）撰写初稿外，其余部分均由笔者撰写并统编和校对全书。

本书是国家社科基金项目《“短语理论”与语料库的“教学加工”》的研究成果之一，曾得到该项目（编号09BYY067）和华南师范大学“十一·五”规划教材项目的资助以及华南师范大学外国语言文化学院师生的试用反馈意见，特此致谢。

何安平

2010年8月于广州

致 谢

本书所附光盘中的文本有些是从国内外的英语语料库中提取的少量检索结果，有些是国内外一些教材和网站上的少量文字材料，以微型文本的形式供本书的工作坊使用，以更好地帮助读者直观地学习语料库的操作；光盘中还包含一些语料库检索工具。这其中多数为免费资源，其他材料有些已经获得授权，有些由于种种原因无法联系到作者，除在书中标注过的以外，在此一并致谢。

BNC: *The British National Corpus*. Lou Burnard, L. 2001. Consortium, UK.

BROWN: *The Brown University Standard Corpus of Present-Day American English*. Kucera, H. & N. Francis. 1960. Brown University, U.S.A.

CHILD: *Children's Literature Corpus*. Budzilowicz, M. New York University, U.S.A.

COLT: *The Bergen Corpus of London Teenage Language*. Stenstrom, A. 1993. Bergen University, Sweden.

FLOB: *The Freiburg-LOB Corpus of British English*. Hundt, M., Sand, A., & R. Rainer Siemund. 1998. Englisches Seminar, Albert-Ludwigs-Universität Freiburg, Germany.

FROWN: *The Freiburg-Brown Corpus of American English*. Mair, C. 1999. Albert-Ludwigs-Universität Freiburg, Germany.

LLC: *The London-Lund Corpus of Spoken English*. Randolph Quirk, R., Greenbaum, S. & J. Svartvik. 1980. University College, UK. & Lund

University, Sweden.

LOB: *The Lancaster-Oslo/Bergen Corpus of British English*. Johansson, S., Leech, G. & H. Goodluck. 1978. Lancaster University, UK/ University of Oslo & Bergen University, Norway.

LOBTAG: *The Tagged LOB Corpus*. Johansson, S., Atwell, E., Garside, R. & G. Leech. 1986. Norwegian Computing Centre for the Humanities Bergen, Sweden.

Willis, J. & D. Willis. 1988. *Collins COBUILD English Course*. London: Collins.

陈琳等. 2006. 普通高中英语课程标准实验教科书 英语 [S]. 北京：外语教学与研究出版社.

桂诗春, 杨慧中. 2003. 中国学习者英语语料库 CLEC [M]. 上海：上海外语教育出版社 .

刘道义等. 2007. 普通高中课程标准实验教科书 英语 [S]. 北京：人民教育出版社.

翟象俊, 郑树棠, 张增健. 1999. 21 世纪大学英语 [S]. 上海：复旦大学出版社.

http://www.antlab.sci.waseda.ac.jp/software.html

http://www.edudown.net/student/English/reading

http://www.fleric.org.cn/pub/GK2008_range2.0.txt

http://www.pep.com.cn/download/gzenglish/yuliao.zip

http://politicalticker.blogs.cnn.com/2010/02/06/obama-promotes-small-businesses

http://sfs.scnu.edu.cn/corpus4u

http://sfs.scnu.edu.cn/corpus4u/show.aspx?id=348&cid=20

http://www.wwenglish.com

目　录

第一章　语料库语言学简介

本章提要：

简要介绍语料库语言学的学科定义、学科背景、发展沿革以及学习如何进入各类语料库样品，对语料的类别和体例进行观察。

第一节　学科定义

语料库语言学 (corpus linguistics) 是当代语言学的一门新兴学科。它运用计算机手段对巨量的语言资源库（又称语料库，corpus 或 corpora）进行高速提取并准确显示批量语言使用的实际情况，从而揭示语言使用的倾向性规律及其所传递的意义、功能乃至思想意识。关于语料库有多种定义，以下是语料库语言学家们对语料库和语料库语言学所作的界定。

文献阅读：

注意阅读以下定义中一些反复出现的词和短语，从中掌握语料库以及语料库语言学的核心内容，包括语料的规模、来源、储存和提取手段以及用途。

1. 关于语料库

- A collection of naturally occurring language text, chosen to characterize a state or variety of a language.
- A body of naturally-occurring (authentic) language data which can be used as a basis for linguistic research.

(Sinclair, 1991: 171)

- In the past thirty-five years, the term corpus has been increasingly applied to a body of language material which exists in electronic form, and which may be processed by computer for various purposes.
  (Leech, 1997: 1)

- A collection of texts, especially if complete and self-contained: the corpus of Anglo-Saxon verse.
- ..., a body of texts, utterances, or other specimens considered more or less representative of a language, and usually stored as an electronic database.
  Currently, computer corpora may store many millions of running words, whose features can be analysed by means of tagging (the addition of identifying and classifying tags to words and other formations) and the use of concordancing programs. Corpus linguistics studies data in any such corpus ...
- Bodies of natural language material (whole texts, samples from texts, or sometimes just unconnected sentences), which are stored in machine-readable form. Computer corpora are rarely haphazard collections of textual material: they are generally assembled with particular purposes in mind, and are often assumed to be (informally speaking) representative of some language or text type.
  (McArthur & McArthur, 1992)

2. 关于语料库语言学

- Corpus linguistics is not an end in itself but is one source of evidence for improving descriptions of the structure and use of languages, and for various applications, including the processing of natural language by machine and understanding how to learn or teach a language.
  (Kennedy, 1998: 1)

- What, then, is a 'corpus linguist'? I would like to think that it is a linguist who tries to understand language, and behind language the mind, by carefully observing extensive natural samples of it and then, with insight and imagination, constructing plausible understandings that encompass and explain those observations. Anyone who is not a corpus linguist in this sense is, in my opinion, missing much that is relevant to the linguistic enterprise.
  (Chafe, 1992: 96)

阅读反思：

从以上论述可以看出，语料库是在一定原则下收集的批量的口头或笔头语篇素材，并且以电子版本的形式储存在电脑中，用于语言的量化调查和质性分析。它是一种事实性的语言资源，用来改进人们对语言结构和语言应用的描述并且应用于多个方面，包括：如何用计算机来加工自然语言，以及如何理解、如何教授语言。语料库语言学的主要活动就是语料库语言学家力图通过仔细观察批量的自然语言样品来理解语言以及语言背后的思想。然后，通过本质观察（insight）和想象来构建对这些观察结果的理解。由此推及语料库语言学的语言观是将语言看作一种社会行为和做事方式，所以要将实际运用中的语言作为研究对象，通过计算机技术辅助来对海量的语言事实进行调查和分析，以得出对语言的本质和运用规律的更深刻和更全面的认识，它反映的是一种经验主义和实证研究的语言哲学思想。

第二节　发展简史

语料库语言学作为语言学研究的一门新兴学科，起源于 20 世纪 60 年代的英美国家。如果从语料库的规模、语料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]：

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（或者是更少），如BROWN和LOB。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1988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)。进入本世纪以来，语料库语言学在国内逐步推广开，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，

1 参考 Renouf, 2006: 28.

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》（何安平和郑旺金，2009）[2]；《商务英语语料库》、《中医英语语料库》等也在建设中。

至于语料库未来的发展方向，则很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。

第三节　语料库类型与体例

语料库的种类有很多，例如笔语语料库和口语语料库、生料语料库和附码语料库以及通用语言语料库和专业语言语料库等。本节学习如何使用语料库检索工具提取并了解各种类型语料库的情况。

工作坊1　观察语料库

教学目的：

通过使用本书附带光盘中的语料库检索软件 AntConc 和语料库样品来学会：

1）进入语料库；

2）阅读语料库中的内容，包括常见的附码或代号；

2　详见 http://sfs.scnu.edu.cn/corpus4u 主页上的教材语料库在线。

3）看看自己能够在 10 分钟内打开多少不同种类的语料库，并且说出各自的一些体例特点；记录下你最感兴趣的两个语料库样品的特征。

教学语料：

本书光盘中 data\chapt1\ 目录下的各种语料库样品，使用工具是本书光盘中的 soft\AntConc3.2.1w[3]。

操作指引：

1）使用 AntConc3.2.1w 的 File View 工具打开本书附带光盘中的各类语料库样品，检索界面见图 1，具体步骤见本书第二章第二节。

2）观察各种体例不同的语料库样品检索结果（见图 1 至图 7），尝试进行分类，如：国内 / 国外，英语本族语者 / 英语学习者，笔语 / 口语，生料 / 附码，普通英语 / 专业英语等。

3）使用 File → Open File(s) 打开一个或多个文件，使用 File → Open Dir 打开一个目录内的全部文件。

4）查看语料库的方法是点击 Corpus Files 栏内的一个文件名，然后点击检索界面上方的 File View → Start，见图 1。

5）退出语料库的方法是点击 Corpus Files 栏内要关闭的某个语料文件名，然后点击 File → Close File；也可以点击 File → Close All Files 来关闭所有检索文件。退出当前检索界面的方法是点击主界面左下角的 Reset 框。

6）关闭检索工具的方法是点击检索界面右上角的 ×。

3 该软件可在互联网上免费下载，自行安装在 PC 机上便可开始检索各类语料库；3.2.0w，3.2.1w，3.2.2w 三个版本的功能大致相同，不影响对本书工作坊的操作。