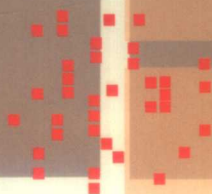


中国英语教师丛书

语料库语言学与英语教学

何安平 著

丛书主编 顾曰国
陈国华
王初明



外语教学与研究出版社

FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

语料库语言学与英语教学

丛书主编 顾曰国
陈国华
王初明

FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

北京 BEIJING

图书在版编目(CIP)数据

语料库语言学与英语教学/何安平著. —北京:外语教学与研究出版社, 2004.9
(中国英语教师丛书)

ISBN 7-5600-4480-8

I. 语… II. 何… III. ①词语—应用—外语教学 ②英语—教学研究
IV. ①H09 ②H319.3

中国版本图书馆 CIP 数据核字(2004)第 099743 号

语料库语言学与英语教学

何安平 著

* * *

责任编辑:唐 辉

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:北京新丰印刷厂

开 本:850×1168 1/32

印 张:6.75

版 次:2004 年 10 月第 1 版 2004 年 10 月第 1 次印刷

书 号:ISBN 7-5600-4480-8/H·2105

定 价:9.90 元

* * *

如有印刷、装订质量问题出版社负责调换

制售盗版必究 举报查实奖励

版权保护办公室举报电话:(010)88817519

总序

外语教学与研究出版社于2000年引进和出版了国外《当代语言学和应用语言学文库》首批54种，次年再次推出58种，对中国读者来说《文库》是名副其实的雪中送炭，所引起的空前的反响足以为证。它不仅仅成为语言学界的一件美谈，也不失为一种文化现象。它的底层是中华学子对知识的渴求和力图洋为中用、科学强国的治学动机。它使人们回想起五四新文化运动时期的情景。被船坚炮利的洋人攻破国门，引发了中西文化史无前例的碰撞。采东、西洋谋自强很快取得共识。问题是采什么？如何采？“西学中源”论，“中学为体，西学为用”论，“用夷变夏”论，等等，成为西学东渐争论的焦点。

以本人较为熟悉的修辞学为例。陈望道把五四新文化时期称为“中外修辞学说竞争时期”。这是因为根据宗廷虎、李金苓（《中国修辞学通史·近现代卷》吉林教育出版社1998），最早引进国外修辞学说的是汤振常的《修词学教科书》（1905）。我们因此可以说，汉语古代修辞学是土生土长的，只有输出，没有输入。输出的典型例子就是唐代来长安留学的日本僧人空海（法号遍照金刚，追封尊号弘法大师，774—835）。他的《文镜秘府论》，“唐人巨言，近在其中”（王利器校注，1983，前言引《半江暇笔》，第11页）。汉语现代修辞学是中外修辞学碰撞后的产物。陈望道认为中国修辞学的独立是以“一九一九年的五四运动来做一个自然的界线”（《修辞学发凡》1962 [1932]:274）。西方的 rhetoric 进入中国后，学者的反应可以归为五类：（1）西体中用（西方的理论原封不动，换上中文的例子）；（2）西法中用（西方的分类法，换上中文的例子）；（3）选择性的引进；（4）批判性的引进；（5）全盘否定。这段时期距今达百年，我们可以说有了足够的时空间来品头论足。直觉的判断是和第（1）和

第(5)的历史沉淀甚少,第(2)、(3)、(4)对后学有启迪和开创之处较多。深究起来这个评语并不公允。因为西体中用在中西文化初期碰撞时起到了引进来的作用,全盘否定则起到了当头棒喝、避免盲目崇拜的作用。然而,这并不等于说采西学的方式无论怎样都是可取的。如对西学一知半解,对国学又浮于皮毛,却同时摆出一副专家的气派在引进西方 rhetoric 时就不乏其例。这是要引以为戒的。

至此,读者也许要问,上面这段议论跟为本书作序似乎有点风马牛不相及。非也。篇首提到的《文库》凡112部,都是西人所著。作为该《文库》的延续,外语教学与研究出版社策划出版另一个系列,宗旨是推出我国学者洋为中用的研究成果。如果说《文库》是栽树施肥的举措,那么新系列则是开花结果的收获。这次推出的有4部,分别是董燕萍的《心理语言学与外语教学》、何安平《语料库语言学与英语教学》、张德禄、苗兴伟、李学宁的《功能语言学与外语教学》和蓝纯的《认知语言学与隐喻研究》。尽管它们的研究主题不同,却都有一个共同的特点,那就是力图中西结合,推陈出新。作为这些著作的第一读者,本人自然有些读后感。上文关于五四新文化时期引进西方 rhetoric 的一段文字,正是本人在读这些著作时所引发的回忆。在此写成文字,分享读者,本人以为作为序言是再好不过的了。读者也许会从中得到启迪,用历史的立体的眼光来审读这个新系列,作出自己的评价。

顾曰国

2004年9月于北京

序

运用语料库，这种汇集了大量真实语言的信息资源来对语言理论、语言结构以及语言的使用进行实证性的研究已经成了近年来语言学发展的一个重要趋势。如今全球用于专门研究语言的语料库就有数百个。尽管当初第一批电脑语料库仅仅是用来研究英语，但是自 21 世纪初以来的语料库已经可以用来研究除英语之外的多种语言，其中也包括汉语。

基于语料库的研究在语言理论、语言描述、自然语言加工处理（包括语音合成和自动翻译）以及外语教学的理论与实践等领域已经作出了重要的贡献，我特别高兴看到何安平博士写成的这本关于语料库语言学与外语教育教学的新书。何博士具有语料库建设和开发研究的经验，我认为她是将语料库语言学传播给更广大中国读者的理想人物。特别是她利用语料库所进行的外语教学课程设计和教材分析等研究在当今尤为及时。我相信《语料库语言学与英语教学》对于所有从事语言研究的人士，尤其是外语教师和攻读语言学的学生都是一本很有价值的书。

新西兰·惠灵顿维多利亚大学

G. 肯尼迪教授

2001 年 10 月

■肯尼迪教授是本书作者的博士导师，欧洲和大洋洲语料库语言学早期的开拓者之一。其撰写的 *Introduction to Corpus Linguistics* 《语料库语言学入门》1998 年由剑桥大学出版社出版并于 2000 年开始在中国发行。

前言

这是一部介绍和探讨语料库语言学应用于外语教育教学的理论与实践的书。语料库语言学是近30年来在国外迅速发展起来的跨领域学科（语言学与计算机科学）。在教学中引入大型电脑语料库是当今外语教育发展的新趋势。目前国外已经出版了多部介绍语料库应用于语言研究与语言教学的专著或论文集，而国内在这方面的理论发展与实际应用似仍在起步阶段。笔者作为曾在国外攻读电脑语料库语言学的博士生并在1997年回国后一直从事语料库建设、开发和教学应用，非常愿意与国内的同行及读者共同学习和分享这一领域的最新发展及丰硕成果，并藉此推动该学科在国内的发展与应用。

本书的撰写遵循两大原则，一是理论与实际相结合，这就是针对国内外外语教育教学的实际需要介绍国外语料库语言学的最新发展及应用成果。二是学习与实践相结合，这就是从解剖大批研究实例入手，着重评述其研究方法和实际操作过程，让学习者在吸取他人研究经验的同时自己也有信心动手实践并有所创新。另外书中多处提供的网页和网址以及书后附录的检索软件使用说明也可以帮助读者自己动手寻找语料库和使用检索工具。

本书在撰写的过程中得到全国高校人文社科重点研究基地广东外语外贸大学外国语言学及应用语言学研究中心的资助和华南师范大学外国语言文化学院的支持和鼓励。国外也有许多专家学者为此书提供了宝贵资料，其中包括 S. Granger 教授（比利时卢汶大学），A. Stenström 教授（挪威伯根大学），A. McEnery 教授（英国兰卡斯特大学），P. Perters 教授（澳大利亚麦考尔大学）等等。作者的博士生导师，新西兰维多利亚大学的 G. Kennedy 教授更是在百忙之中亲笔为本书作序。在此，作者向他们表示衷心的感谢并期待各位专家同行批评指正。

PREFACE

Professor Graeme Kennedy
Victoria University of Wellington
New Zealand

The use of large collections of texts, known as corpora, as sources of evidence about the structure and use of languages has been one of the major developments in linguistics in recent years. There are now hundreds of collections of texts in electronic form which have been compiled for research purposes. Although the first modern electronic corpora were compiled for the study of English, corpora are now also available, at the beginning of the 21st century, for the study of many other languages including Chinese.

Corpus-based research has already been responsible for significant developments in linguistic theory, linguistic description, natural language processing (including speech synthesis and automatic translation), and in foreign language teaching theory and practice. It is a great pleasure to welcome the completion of Dr. He Anping's new book on the possible contribution of corpus linguistics to foreign language education. Her experience in corpus compilation and corpus analysis had made her an ideal person to introduce corpus linguistic methodology to a wider audience in China. In particular, Dr. He's commitment to the improvement of foreign language education through the use of corpora to influence curriculum design and classroom teaching materials has been especially timely. I am confident that *Corpus Linguistics and English Language Education* will be of value for all scholars interested in the study of language and especially for language teachers and students of linguistics.

目 录

第一章 语料库语言学简介	1
1.1 语料库语言学的发展历程	1
1.2 语料库的类型	3
1.3 语料库的检索工具	6
第二章 语料库语言学的研究层面与方法	9
2.1 语言结构层面	10
2.1.1 辨析近义词	10
2.1.2 调查短语结构的搭配模式	11
2.1.3 探讨意念的表达形式	15
2.2 语篇结构层面	17
2.2.1 构建口语篇章结构词的分析模式	17
2.2.2 构建篇章类型的语言特征模式	19
2.2.3 分析语音语调的语篇功能	22
2.3 话语行为层面	25
2.3.1 对对话行为的综合描述	25
2.3.2 对医生-病人会话的语用分析	30
2.4 文学作品分析	34
2.4.1 诗歌作品的赏析	34
2.4.2 两种文体风格的对比	38
2.5 研究方法小结	42
第三章 语料库与英语教育教学	45
3.1 语料库与教学资源	45
3.2 语料库与教学大纲设计及教材编写	47

3.2.1 利用语料库信息检验教学大纲的语法排序	48
3.2.2 利用语料库信息编排教学内容	49
3.2.3 利用语料库信息调整教学内容的重点	50
3.3 语料库直接用作课堂教材	51
3.3.1 从报刊语料库辑选阅读资料	51
3.3.2 归纳新闻题材的语篇框架和修辞特征	52
3.3.3 通过报刊语料库学习词汇	53
3.4 基于语料库开发多媒体教学课件	53

第四章 语料库与课堂教学 57

4.1 语料库应用于课堂教学的理论基础	57
4.2 语料库与词汇教学	59
4.2.1 堂上即席制作的语境共现	59
4.2.2 学习叙事口语中的常见词	62
4.2.3 基于语境共现设计的词汇学习活动	64
4.2.4 利用语料库设计词汇测试	78
4.2.5 小结	80
4.3 语料库与语法教学	81
4.3.1 基于语料库的语法教学设计	81
4.3.2 小结	90
4.4 语料库与语音教学	90
4.4.1 学习划分英语语调单位	91
4.4.2 学习重音和语调模式	91
4.4.3 利用真实语料解决学习者的语音问题	92
4.5 语料库与文体学教学	93
4.6 语料库与阅读分析课教学	96

第五章 语料库与教学研究 101

5.1 语料库应用于教学研究的理论基础	101
5.2 基于语料库的教材研究	102
5.2.1 对不同教材的词汇对比分析	102
5.2.2 利用语料库分析教材的话题分布	104
5.2.3 利用语料库探讨语法内容的排序	106
5.2.4 利用语料库评估练习的类型与层次	107

5.3 基于语料库的课堂话语研究	108
5.3.1 从话轮角度分析课堂师生会话	109
5.3.2 对教师课堂提问话语的分析	110
5.3.3 对课堂中磋商式话语语段的分析	112
5.4 基于学习者语料库的中介语研究	115
5.4.1 语料库对中介语研究的意义	115
5.4.2 学习者的英语差错类型调查	117
5.4.3 分析学习者对英语连接语的使用	120
5.4.4 开发英语学习工具软件	124
结语 语料库语言学发展的前景	127
参考文献	129
附录 A 国外部分语料库的内容结构及附码标记说明	141
附录 B 国外部分语料库检索工具的使用说明	179

第一章

语料库语言学简介

现代语料库语言学 (modern corpus linguistics) 是20世纪中后期兴起的一门语言研究科学。在这之前对语言研究的方法和流派已有多种多样。其中一种被称为“沙发里的语言学 (armchair linguistics)”, 指的是学者们坐在办公室里凭借自己头脑中的语言直觉和智慧, 通过主观臆想来撰写语言的规范。另外一种被称为“诊疗室里的语言学 (therapy linguistics)”, 这是基于个别或少数失语症患者的话语语料来研究患有语病的语言模式, 它关注的是研究者能够利用语言来做什么而不是一般人如何使用语言。还有一种就是“语料库里的语言学 (corpus linguistics)”。这是基于大量真实的语言资料, 从调查语言信息的分布频率入手来研究语言在现实使用中的规律和模式。基于真实语料来研究语言并不是近一两个世纪才发生的事情, 然而, 如今被人们称作“现代语料库语言学 (modern corpus linguistics)”的则是指专门对大规模储存于计算机里的语料库进行研究的学问。从这个意义上讲, 电脑语料库是一个由大量在真实情况下使用的语言信息集成的、可供计算机检索的、专门作研究使用的巨型资料库。它以其容量大、语料真实、检索快捷准确等独特的优势在现代语言研究和语言教育中正发挥着越来越重要的作用。语料库语言学不仅仅是研究方法论的一个重要突破, 它更孕育着对语言描述框架乃至语言观念的不断更新 (何安平 1997:21-26)。

1.1 语料库语言学的发展历程

运用语料库进行语言研究可以追溯到19世纪末,

但当时的研究手段还只停留在卡片制作和人工检索的阶段,其成果也仅用作编撰语法书或词典的参考。1890年德国的 Kaeding 为了收集德语的字母和单词频率信息以发展德语速记模式,曾动用了 5,000 多名助手花了数年时间处理了 1,100 多万词次的口、笔头语料(转引自 Kennedy 1998: 25)。20 世纪初,美国教育家 Thorndike (1921) 为了改进英语教育教学而收集了 4,500 万词次的英语语料,其中有 3/4 的语料是来源于《圣经》和其他经典文学作品。通过对这些书面语料的词频调查编写出英语基本词汇表及词典等等。在电脑语料库问世之前,对真实语料的收集和研究不仅在数量、规模和代表性方面很受局限,而且检索起来也很费力费时。20 世纪 50 年代至 60 年代,英国知名学者 R. Quirk, G. Leech 和 S. Greenbaum 为了描述英语语法而开始合作建立百万词次的英语用法普查语料 Survey of English Usage (SEU) (Quirk 1968)。先是纸质语料,后来在 80 年代初建成电脑版本。这是从传统语料库过渡到电脑语料库的重要标志。而将大量语料作电脑化处理 (computerized corpus), 即用电脑进行注释加工、电脑解码以及电脑检索、取样和统计分析等等则是近 30 年来的事。20 世纪 60 年代至 70 年代世界上最先建立的 *Brown University Standard Corpus of Present-Day American English Corpus (BROWN)* (Francis & Kučera 1964) 和 *Lancaster-Oslo-Bergen Corpus (LOB)* (Johansson et al 1978) 就分别汇集了 100 万词次的美国英语和英国英语,其内容覆盖了新闻、宗教、技术、娱乐、小说、信函等 15 种书面语体。接着在 80 年代建成的 *London-Lund Corpus of Spoken English (LLC)* (Svartvik & Quirk 1980) 则是第一部汇集了 50 万词次的英语口语语料库,其中有正式或非正式的个人演说、评论、双人或多人会话、讨论、采访和电话录音等等。这 3 大经典语料库为现代语料库语言学奠定了坚实的基础,因为它们所采用的不是通过臆想或杜撰的语言,也不是仅仅从某些著作抄录下来的若干例句,而是有目的有系统地收集的大量在现实生活中使用的书面语和口头语;并使用先进的电脑科技手段进行储存和检索,从此将语言研究建立在一个博大的、真实可靠的语料基础上。

20 世纪 80 年代至 90 年代是语料库语言学发展的第二个阶段,其标志之一是世界各地都开始建设自己的语料库并且开始跨国联合建立国际性的语料库。我国首批建成的 50 万词次的广州石油大学石油英语语料库 *GPEC* (Zhu Qibo 1989) 和 100 万词次上海交通大学科技英语语料库 *JDEST* (Yang Huizhong 1986) 也是在这个时期。在国外有代表性的是 90 年代集有 1 亿词次的英国国家语料库 *British National Corpus (BNC)*, 它含有 9,000 万词

次的书面语和 1,000 万词次的口头语。还有由 Greenbaum (1991) 主持建立的国际英语语料库 *International Corpus of English (ICE)*。它汇集了全球 20 多个国家和地区的英语语料, 单是每个子语料库就已经收有书面和口头语料各 50 万词次。

20 世纪 90 年代中期以来是语料库语言学发展的第三阶段。其发展趋势主要表现在以下 3 个方面: (1) 大规模、多品种语料库的建设; (2) 对语料库的深加工研究; (3) 语料库在与语言相关的各个领域中的广泛应用。由于计算机等电子传播和储存技术的快速发展带动了语料库的建设规模迅速扩大, 二十世纪六七十年代建立的第一代电脑语料库, 如拥有百万词次的 *BROWN* 和 *LOB* 等如今看来已算不上大型语料库, 因为 90 年代建立的 *BNC* 就已经达到 1 亿词次, 而 21 世纪初已实现通过因特网检索的在线语料库 (*WEBCORP*) 更是达到了 10~50 亿词次, 平均每天更新的新闻语料网页就达 200 万 (参考 Pacey M. & A. Renouf 2001)。可见, 建立大型语料库已经不再是很困难的事, 接下来的是如何建立各种有地域、语体、乃至语篇特色的大型语料库以及如何对现有的语料库进行深层次的开发与研究的问题。例如, 对原始语料作多层面信息的自动附码处理; 对多功能语料库检索工具的开发研制; 对自然语言的自动识别和电脑化处理系统研究等等 (详见何安平 2001c)。欧洲语料库语言学的一位开拓者 J. Sinclair (1991:1) 曾这样总结 30 多年来语料库语言学发展所带来的观念转变。他说, 处理数百万词次的语料在 30 年前被认为根本不可能; 在 20 年前被认为有一丝可能但难以想象; 在 10 年前被认为很有可能但仍然困难重重; 而到了今天, 这一切都已变得十分平常和普遍。

1.2 语料库的类型

建立电脑语料库可以在全世界、全国或某地区某单位乃至个人的能力范围内实施。有些语料库是属于国家的语言资源档案库; 有些语料库在解决了语料版权问题之后可以成为流通的文化商品, 但是更多的语料库是用于单位和个人的学术研究。现时国内外流行的电脑语料库类型可以包括:

- 原始语料库 (Raw corpora), 这就是将现实中使用的口头和笔头语用文字形式收集起来, 按一定原则 (语域, 语体, 历时, 共时等) 归类汇编起来的各种语料库。例如 1999 年挪威 Bergen 大学新版的 *ICAME Collection of English Language Corpora* 光盘就包含了世界各地的英语口语和书面语, 成人语和青少年语, 当代语和古代语等 21 个语料库共 1,700 万词

次。其中有 90 年代后期建立的可与先前的 *BROWN* 和 *LOB* 作历时性对比的新版 *FLOWN* 和 *FLOB* 以及 *WordSmith* 和 *TACT* 等一批检索工具 (详见何安平 2001a:75-78)。

- 附码语料库 (Annotated corpora), 这是指对原始语料进行了词性、语法、语音、语义或语篇乃至语用标记附码的语料库。例如, 已作词性附码的 *BROWN* 和 *LOB*, 已作语法附码的 *ICE-GB* (国际英语语料库英国子语料库) 和已作语音韵律附码的 *LLC* 和 *SEC* (英语口语语料库) 等等。
- 平行语料库 (Parallel corpora), 这是指两种或多种语言在句子乃至单词短语层面上实现同步对译的互动语料库, 如英语与法、德、西班牙等语种的平行语料库 *CRATER* (McEnery & Oakes 1996) 和英汉双语平行语料库 (中国外语教学研究中心基地 2000) 等。
- 学习者语料库 (Learners corpora), 即非母语学习者的口头和书面语料库, 其中包括注有学习者拼写和语法差错标记以及改错提示的语料库。如, *ICLE* (国际英语学习者书面语料库), *LINDSEI* (国际英语学习者口语语料库) (Granger 2000) 和 *CLEC* (中国英语学习者书面语料库) (桂诗春 2001) 等等。
- 网格式语料库 (Lattice corpora), 这是指对自然语言 (包括口语和手写字) 进行自动语音和手写体识别处理之后生成的语料库 (Atwell 1996)。

归总起来, 语料库可分为两大类: 原始语料库 (未经任何加工的) 和附码语料库 (经过整理加工并附加上各种符号标记的)。其中附码语料库又可以再分出若干类, 以下是其中几例的样品及说明:

A. 附有词性标注的语料库 (lexically tagged corpus), 如:

A01 16 ^ he_PP3A believes_VBZ that_CS the_ATI House_NPL of_IN Lords_NPTS
A01 16 should_MD be_BE abolished_VBN and_CC that_CS
A01 17 labour_NN should_MD not_XNOT take_VB any_DTI steps_NNS which_WDTR

这是从已经作了词性标注的 *LOB* (英国书面语语料库) 中 Section A (即第一部分: 新闻报道) 节录的一个片段, 其原始语料文字是: ...he believes that the House of Lords should be abolished and that labour should not take any steps which...。其行首的 A01 表示语料库中 Section A 部分的第一号样品。16, 17 代表语句排行的序号。^ 表示一个句子的开始。每个词

_ 之后的大写字母都是词性的代码, 如 PP3A 表示代词第三人称单数, VBZ 表示动词第三人称单数, CS 表示从属连词, ATI 表示定冠词, NPL 表示专有名词, IN 表示介词等等。(详见 S. Johansson 1986:141-149)

B. 附有句法成分标签的语料库 (syntactically parsed corpus), 如:

[S[N Bob_NP N][V likes_VBZ V][N toys_NNS N]_. S]

这是根据 *LOB Corpus Treebank* (详见 Leech & Garside 1991) 的句法附码系统标注的一个简单句。原始语料的句子是: Bob likes toys。附码之后, [S S] 表示一个句子的起止, [N N], [V V] 和 [N N] 分别表示句中的主、谓、宾各种成分的起止。各单词 _ 之后的大写字母是词性代码。_. 表示英语句号。

C. 附有音韵特征标注的语料库 (prosodically annotated corpus), 如:

```
1 1 1 10 1 1 B 1 1 ((of ^Spanish)) . graphology# /
1 1 1 20 1 1 A 1 1 ^w=el# /
1 1 1 30 1 1 A 1 1 ((if)) did ^y/ou _set _that# - /
1 1 1 40 1 1 B 1 1 ^well |J/oe and _I# /
1 1 1 50 1 1 B 1 1 ^set it between _us# /
```

这是摘自 *LLC* (英国口语语料库) 的一个片段。对话原文是:

B: ...of Spanish graphology

A: well if did you set that

B: well Joe and I set it between us ...

经过语调和韵律的标注之后, 1 1 1 表示第一类语料 (即日常会话类) 的第一号样品的第一段对话语料。10-50 代表语调组的序号。1 1 表示录音记录的行号。A 和 B 分别代表两个说话者。1 1: 前面的 1 表示录音记录的行号已完结, 未完结的则用 2; 后面的 1 表示该语调组的调头重音 (onset) 的数量, 若无调头重音的则用 0。行文中注有各种调韵特征标记。例如, ^ 和 ' 分别表示调头重音和句重音; \, /, ^, V, 和 = 分别代表降调, 升调, 升降调, 降升调和平调; . 和 - 分别表示短暂停顿和长约一拍的停顿等等

(详见 J. Svartvik & R. Quirk 1980:21-34)。

D. 附有差错类型标记的语料库 (error-tagged learners' corpus), 如:

... Xiao Li, a friend of my [pr4,3-0] liked cats very much. He had two cats. One was very big and the other was quite small...

这是摘自 CLEC (中国英语学习者书面语料库) 的中学生作文语料片段。
[pr4, 3-0] 表示前面的 my 是人称代词所有格类型的错误, 其中 3-0 表示 my 涉及差错上下文语境是从 my 之前的第三个词起 a friend of my (详见桂诗春 1998)。

1.3 语料库的检索工具

建成大型语料库仅仅是完成了语料的收集、整理和加工。要进行语料库的调查研究还必须依靠各种电脑检索的系统工具。当今除了比较流行的 TACT¹ 和 WordSmith² 之外, 还有许多针对专门语料库设计的检索软件, 如专作单词类别和频率分析的 VocabProfile³, 专作口语语篇内容分析的 ACAMRIT⁴ 以及用于句法句型检索分析的 ICECUP3.1⁵ 等。这里介绍的仅仅是一套比较适合英语教育和课堂教学的检索系统软件 MicroConcord (Scott M. & T. Johns 1995)。

MicroConcord (简称 MCO) 是 1995 年由英国牛津大学计算机中心的 Mike Scott 和 Tim Johns 共同研制的一个具有 30 多种功能, 操作简便的语料库检索软件。它的操作指令为简单英语, 可在 DOS 3.0 或 Windows 95 以上的环境下运行。主要优点是它可以检索任何的纯文本或 Word 类型的英语文本。它可以瞬间在数百万词次的语料库中检索某一单词、短语、句型及任何符号的出现频率及上下文语境, 并且可以当场对检索结果 (output file) 进行统计、分类、删节、归档、储存等编辑工作或隐去检索项作编写教材或练习之用。因此特别适合一般英语教师和学生用于教学研究和英语学习 (详见本书第五章)。以下是它的 3 大主要功能介绍 (详细使用见本书的附录)。

a) 检索功能

- 直接检索一个或多个词, 短语及限定其大小写形式, 检索项简称 SW。
- 使用通配符 (wildcards) 如 * 和 ? 检索一个词的局部或一个短语的局部。
- 使用语境搭配词 (context word) 提取 SW 之前 9 个词及之后 9 个词范围