

教育部人文社科规划基金项目 (12YJA740060)

华中科技大学外国语学院学术专著出版专项基金

Corpus-based Language Research and Pedagogical Applications

基于语料库的 语言研究与教学应用

潘 璠 / 著



中国社会科学出版社

教育部人文社科规划基金项目 (12YJA740060)

华中科技大学外国语学院学术专著出版专项基金

资助

Corpus-based Language Research and Pedagogical Applications

基于语料库的 语言研究与教学应用

潘 璠 / 著



中国社会科学出版社

图书在版编目(CIP)数据

基于语料库的语言研究与教学应用 / 潘璠著. —北京: 中国社会科学出版社, 2012. 9

ISBN 978 - 7 - 5161 - 1487 - 2

I. ①基… II. ①潘… III. ①语料库—应用—语言教学—教学研究
IV. ①H09

中国版本图书馆 CIP 数据核字 (2012) 第 228648 号

出 版 人 赵剑英
责任编辑 夏 侠
责任校对 徐 楠
责任印制 李 建

出 版 中国社会科学出版社
社 址 北京鼓楼西大街甲 158 号 (邮编 100720)
网 址 <http://www.csspw.cn>
中文域名: 中国社科网 010 - 64070619
发 行 部 010 - 84083685
门 市 部 010 - 84029450
经 销 新华书店及其他书店

印 刷 北京大兴区新魏印刷厂
装 订 廊坊市广阳区广增装订厂
版 次 2012 年 9 月第 1 版
印 次 2012 年 9 月第 1 次印刷

开 本 710 × 1000 1/16
印 张 15.75
插 页 2
字 数 246 千字
定 价 47.00 元

凡购买中国社会科学出版社图书, 如有质量问题请与本社联系调换

电话: 010 - 64009791

版权所有 侵权必究

序 言

潘璠老师的《基于语料库的语言研究与教学应用》即将付梓，倍感喜悦。

初识潘璠是在 2003 年的上海。当时，我在上海交通大学举办一个语料库语言学的国际会议。潘璠宣读了一篇关于研究生中介语分析的论文，颇有新意，遂与其切磋、相识。后来，我受命组稿会议论文专辑，潘璠的这篇论文也入选其中，有幸发表于 2004 年第二期的《现代外语》。10 年中我不时漂泊，京沪穿梭，北国南疆。但无论何时何地，每每获悉她的进展，读到她的新作，都倍受鼓舞。

潘璠是国内很早就进入语料库研究的青年学者。某种意义上讲，她和中国的语料库语言学一起成长、发展、提升。她的硕士生导师冯跃进老师是我留英时伯明翰大学的校友。20 世纪 80 年代末 90 年代初，语料库语言学已在欧洲学界如火如荼，伯明翰大学俨然成为国际语料库语言学最重要的思想与方法发源地，国内则刚刚起步。90 年代中期，潘璠已开始跟随冯跃进老师开始了语料库建设与研究，并着手向国内学界引介国外语料库研究的概略。自那时以来，她一直默默坚守着武汉的这块语料库研究阵地。无论教学任务有多繁重，无论研究资源是何等匮乏，都未改变她对学科的忠诚与事业的信念。都市的喧嚣，人心的浮躁，未能改变她的本初个性：低调为人，潜心为学，避虚求实，一步一个脚印地扎实前行，换来一个又一个的厚重收获。

如书名所示，这是一本紧密联系外语教学实际的语料库语言学专著。毋庸置疑，语料库语言学从开创之初就与应用语言学或语言教学结下了天然联系，而中国国内的语料库研究更是与外语教学研究密不可

分。潘璠老师的这本书涉及这个领域多层面的问题探索，从有关的理论评述和方法梳理，到学习者语料库建设及中介语对比分析，乃至语料库在外语教学中的应用，皆入其中。全书探讨了广泛的经典话题：词语搭配、语义韵、同义词辨析、学习策略研究、语体研究、教材编写、数据驱动学习，林林总总且体系严密。这些研究基于作者对语言事实的细致观察，对大量数据的缜密准确处理，对相关理论问题的严谨审慎概括，源于她对诸多学理问题的认真思索，是她多年勤勉治学、辛苦劳作的结晶。全书从一个侧面反映了学习者语言发展及外语教学过程中的现实问题以及可行的应对策略，展现了作者观察问题的特有视角，给我们提供了值得深思和进一步探究的启示与见解。

如书中所言，今日中国的语料库语言学取得了长足发展，绝非 20 世纪 90 年代所可比拟。然而，会聚在语料库语言学旗帜下的是一支庞大的研究队伍，来自不同的学科背景，怀有不同的价值取向；繁荣与隐忧俱在。当一个学科成为众人追随的时髦时，它就多了几分浮躁，多了一些喧闹；核心理念可能被稀释，本质的方法可能被丢弃。比如，我们还要不要坚持对真实语言事实进行实证研究？我们还要不要坚持自下而上的基础研究方略？我们的数据能承载多少命题假设和推理、解释？甘于寂寞，远离浮躁的学人会不屑人云亦云，坚守应有的操守和信念，冷静地将研究持续推进，如潘璠。

外语教学与研究呼唤更多的语料库语言学真品力作！

语料库语言学呼唤更多的严谨缜密探索！

是为序。

卫乃兴

2012 年 8 月 12 日于香港城市大学

前 言

近年来,语料库语言学已逐渐从原来的语言研究辅助工具的边沿地位,一跃而成为语言研究的主流。计算机软件技术的飞速发展使语料库语言学的研究方法更趋多样化,并使其逐渐演变成为一门多学科、多层次、多功能、多角度分析语言现象的学科。在国际上,语料库语言学的影响已逐步渗透到语言研究的各个领域,使语料库成为词典编纂者和语言研究者必不可少的工具。近 10 年来,国内基于语料库的研究也经历了飞速的发展,与 20 世纪 90 年代语料库语言学初引入时完全不可同日而语。但同时,许多外语教师仍然对语料库缺乏了解,不知道如何利用语料库工具开展语言研究和利用语料库信息改进教学。

本书着眼于推广语料库的研究和应用,使之成为广大外语教师和研究者进行教学和研究的重要资源和工具,并方便教师和研究者充分利用语料库资源和语料库研究成果。一方面,通过对大量国内外语料库研究文献的梳理和提炼,使读者对相关领域的理论发展、研究方法和应用实例有更清晰的认识;另一方面,又通过提供一些研究实例供读者参考,在实践方法上予以引导。本书中探讨的一些领域如语义韵理论发展和相关重要概念之间的复杂关系、语料库信息在词典编撰和教材编写中的具体应用、学习者语料库的研究发现、研究不足以及方法论问题、基于多模态语料库开展数据驱动学习等,在国内文献中还未见深入探讨。

本书是一部探讨运用语料库进行语言研究和教学应用的书,共分为三个部分,分别为理论篇、实证篇和应用篇。理论篇为实证篇提供了研究的相关理论和研究方法,实证篇从不同研究层面对学习者的语言特征进行考察,为语料库的教学应用提供指向。三个部分从研究理论到研究

实践再到教学应用,既自成一体,又彼此关联。

第一部分“语料库应用于外语教学与研究的理论探讨”共分三章,探讨的内容包括:学习者语料库的研究方法、研究发现以及方法论问题;基于语域差异、搭配和语义韵等语料库研究方法的同义词辨析;语义韵理论的发展沿革、语用角度的语义韵冲突以及语义韵的潜在应用。

第二部分“基于语料库的学习者语言特征研究”共分五章,按照从易到难的顺序分别从词汇、搭配和语义韵、语义关系、语用和语体五个不同研究层面展开调查。这五个典型性研究案例都涉及学习者语料库研究中的中介语对比分析方法,但在研究方法上又各具特色。这些实证研究包括:应用卡方检验的方法对不同层次学习者的考试回避现象进行对比调查;运用词语搭配研究的方法调查学习者的同义词搭配行为和语义韵倾向;采用定量和定性相结合的方法考察学习者的连接词使用频率和语义关系分布;对比中外论文作者立场副词的使用特征和语用差异;采用多维度分析法考察不同层次学习者的书面语体特征。这些研究既发现学习者语言具有一定的地道性,也发现学习者与本族语者的语言使用存在较大差异,反映出外语教学在教学思想上、教材编写和教学手段上都需要改进。

第三部分“语料库在外语教学中的直接和间接应用”共分三章,第九章和第十章属于间接应用,第十一章为直接应用。前两章分别讨论语料库语言学对外语教学的启示以及如何利用语料库工具和资源改进教材。最后一章则以欧洲多模态教学语料库 SACODEYL 为例,介绍如何基于多模态教学语料库开展数据驱动学习。

本书注重提供大量对教师极为有用的信息。例如,书中介绍了一些基于语料库编写的国外知名词典和教材,以及这些词典和教材如何利用和展示语料库提供的信息。在附录中还提供了一些实用性信息,如多维度分析法涵盖的 67 项语言特征、国内外学习者语料库汇总、学习者语料库的错误标注集、分别与培养听、说、读、写能力相关的各类国外基于语料库的教材,以及常用语料库与检索工具等。

在本书的准备和撰写过程中,笔者得到了国内外多位专家学者的指导和帮助。2009—2010 年在美国北亚利桑那大学语料库研究中心访学期间,得到知名语料库专家 Douglas Biber 教授和 Randi Reppen 教授的悉

心指导。他们为本书提供了大量资料，并为书中第八章的语料赋码和数据分析提供了技术指导。我的导师冯跃进先生在本书的选题和资料准备过程中给予了很多帮助。卫乃兴教授对本书的框架、书中的理论、研究方法和文字表述提出了许多中肯的建议和修改意见。书中部分观点和数据处理方式得益于与濮建忠教授、梁茂成教授和许家金博士的交流，在此一并表示诚挚的谢意。

本书系教育部人文社科规划基金项目“基于 MD/MF 的中外期刊论文语体特征对比研究”（12YJA740060）的研究成果。在出版过程中，得到华中科技大学外国语学院学术专著出版专项基金的资助，特表感谢。

对于书中存在的疏漏之处，敬请广大读者和同行指正。

潘 璠

2012 年 8 月

目 录

序言	(1)
----------	-----

前言	(1)
----------	-----

第一部分 语料库应用于外语教学与研究的理论探讨

第一章 学习者语料库与外语教学研究	(3)
第一节 学习者语料库的发展和研究综述	(3)
一 学习者语料库的建设和发展特点	(3)
二 基于学习者语料库的研究发展	(5)
第二节 学习者语料库研究方法和实例	(6)
一 学习者中介语与本族语之间的对比	(6)
二 不同学习者中介语之间的对比	(7)
第三节 学习者语言特征总结和原因解释	(9)
一 学习者语言特征总结	(9)
二 原因解释	(13)
第四节 学习者语料库在教学中的应用	(17)
一 学习者语料库在词典编撰中的应用	(17)
二 学习者语料库在教材编写中的应用	(20)
三 学习者语料库在计算机辅助教学中的应用	(24)
第五节 学习者语料库研究中的不足和方法论问题	(25)
一 学习者语料库研究中的不足	(25)

二	学习者语料库研究中的方法论问题	(26)
第六节	结束语	(27)
第二章	基于语料库的同义词差异性特征调查	(28)
第一节	研究背景	(28)
第二节	传统词语辨析和基于语料库的词语辨析	(29)
第三节	不同语域词频分布差异性特征	(31)
一	<i>real, actual, virtual</i>	(32)
二	<i>certain, sure, definite</i>	(32)
三	<i>big, large, little, small</i>	(33)
第四节	搭配与差异性特征调查	(34)
一	搭配与语义差异	(34)
二	搭配与语义韵差异	(36)
第五节	句中位置与语义差异	(39)
第六节	结束语	(40)
第三章	语义韵的语料库调查及应用	(43)
第一节	语义韵的定义和实质	(43)
第二节	语义韵研究回顾	(45)
第三节	搭配 类联接 语义倾向 语义韵	(49)
第四节	基于语料库的语义韵研究和语义韵特性	(52)
一	语义韵的语料库研究方法	(52)
二	语义韵的概率属性和文体关联	(54)
三	消极语义韵的主导地位	(55)
四	语义韵与类联接的关联	(57)
第五节	语用视角的语义韵	(58)
一	语义韵冲突(一)——非诚意性的信号	(58)
二	语义韵冲突(二)——反讽的出现	(60)
第六节	语义韵研究的启示和潜在应用	(61)
一	词典编撰	(61)
二	语言研究	(63)

三 英语教学	(64)
第七节 结束语	(65)

第二部分 基于语料库的学习者语言特征研究

第四章 不同层次学习者的应试回避倾向的对比调查	(69)
第一节 回避的研究回顾	(69)
第二节 研究方法	(71)
第三节 研究结果和讨论	(72)
一 修饰副词的使用	(72)
二 时间频度副词的使用	(74)
三 两种副词使用频数的三库排序比较	(76)
第四节 结束语	(78)
第五章 学习者的同义词搭配行为和语义韵调查	(79)
第一节 同义词的搭配行为和语义韵差异研究	(79)
第二节 研究设计	(81)
第三节 研究结果与分析	(82)
一 <i>cause</i> 和 <i>lead to</i> 语义韵分布的两库比较	(82)
二 <i>cause</i> 显著搭配词的两库比较	(83)
三 <i>lead to</i> 显著搭配词的两库比较	(92)
第四节 讨论和启示	(94)
第六章 学习者书面语中连接词用法的语料库调查	(97)
第一节 研究背景	(97)
第二节 连接词的研究回顾	(98)
第三节 研究设计	(100)
第四节 连接词的频率比较	(102)
第五节 语义关系的比较	(108)
一 <i>And</i> 的语义关系分类	(108)
二 <i>And</i> 语义关系的分布比较	(108)

第六节 结束语	(110)
第七章 语用视角下的中外学术论文立场副词对比研究	(112)
第一节 研究背景	(112)
第二节 评价系统和立场副词研究回顾	(113)
第三节 研究方法	(114)
一 语料库和工具	(114)
二 立场副词的分类框架	(114)
三 研究问题	(115)
第四节 研究结果与讨论	(116)
一 立场副词的频数分布比较	(116)
二 中国作者的立场副词使用特征	(119)
三 立场副词的位置分布和语用差异	(122)
第五节 结束语	(126)
第八章 中国学习者书面语体多特征多维度调查	(128)
第一节 研究回顾	(128)
第二节 研究设计	(130)
一 研究工具和语料库	(130)
二 多特征多维度分析法	(131)
三 研究问题	(132)
第三节 数据分析和讨论	(132)
一 交互性维度的语体特征分析	(133)
二 叙事性维度的语体特征分析	(136)
三 指代明确性维度的语体特征分析	(139)
四 说服力维度的语体特征分析	(141)
五 抽象性维度的语体特征分析	(141)
第四节 各语体内部差异分析	(144)
第五节 总结和启示	(145)

第三部分 语料库在外语教学中的直接和间接应用

第九章 语料库语言学对外语教学的启示	(149)
第一节 语料库语言学的优势	(149)
第二节 语料库揭示的语言不对称性分布特征	(150)
第三节 语料库语言学对外语教学的启示 (一)	(153)
一 外语教学中对语言特征的处理现状	(153)
二 对外语教学的启示 (一): 重视典型性语言特征	(155)
第四节 语料库语言学对外语教学的启示 (二)	(156)
一 母语—目标语差异对教学的影响	(156)
二 对外语教学的启示 (二): 兼顾母语—目标语差异	(157)
第五节 结束语	(159)
第十章 语料库研究与外语教材编写	(160)
第一节 语料库研究揭示的外语教材不足	(160)
一 典型和非典型语言特征处置不当	(160)
二 忽略口语的语言特征	(165)
三 未使用真实的语言实例	(167)
第二节 语料库作为教材编写的工具和资源	(168)
一 利用语料库信息编写教材	(168)
二 利用语料库工具编排教学内容	(169)
三 利用语料库信息检验教材内容和排序	(170)
四 语料库作为教学资料来源	(170)
第三节 基于语料库的国外语法书和教材: 实例分析	(171)
一 Longman Grammar of Spoken and Written English	(171)
二 Real Grammar: A Corpus-based Approach to English	(173)
三 Basic Vocabulary in Use	(175)
四 Touchstone	(176)
第四节 结束语	(176)

第十一章 基于多模态语料库的数据驱动学习	(178)
第一节 背景	(178)
第二节 SACODEYL 的构成和建库理念	(179)
第三节 语料库工具、检索类型和方法	(181)
第四节 教学资源和教学活动创建	(184)
第五节 结束语	(187)
参考文献	(188)
附录 1 各国学习者语料库汇总	(213)
附录 2 学习者语料库的失误标注码	(216)
附录 3 各语义范畴的立场副词频数和比例	(217)
附录 4 各语义范畴立场副词的句中位置分布和比例	(220)
附录 5 Biber (1988) 维度中语言特征列表	(222)
附录 6 any 的索引行	(225)
附录 7 语料库与教学大纲	(227)
附录 8 国外知名出版社出版的基于语料库的教材	(229)
附录 9 语料库与检索工具	(231)
附录 10 关键词索引	(236)

第一部分

语料库应用于外语教学 与研究的理论探讨

第一章

学习者语料库与外语教学研究

本章对学习者的语料库的建设现状、研究发展、主要研究方法、研究发现问题根源进行了全面的梳理和分析,讨论了学习者语料库在词典编撰、教材编写和计算机辅助教学中的应用,剖析了学习者语料库研究中存在的不足,以及研究的方法论问题。书中第二部分“基于语料库的学习者语言特征研究”都属于学习者语料库研究,也都涉及本章中讨论的中介语对比分析方法,研究结果分别在不同层面上与本章总结的学习者研究发现相吻合。

第一节 学习者语料库的发展和研究综述

语料库技术的发展使对大量的学习者语料进行量化分析成为可能,正如 Leech (1998) 在前言中所说,学习者语料库的出现为发展中的第二语言习得研究另辟蹊径,提供了更为客观和翔实的数据以及经验性研究的手段和方法。基于语料库的学习者中介语分析能帮助语言研究者观察到学习者学习一门新语言的过程和规律,因而给外语教学和研究带来了崭新的改变,使中介语分析由过去的“主观分析”和“定性分析”为主逐渐向“数据驱动型分析”和“定性定量结合型分析”演变。

一 学习者语料库的建设和发展特点

学习者语料库是通过收集语言学习者各种书面语和口语的自然语料而建立起来的一种学习者语言数据库,也称机读学习者语料库 (Computer Learner Corpus) (杨惠中 2002)。随着计算机技术的飞速发展,从