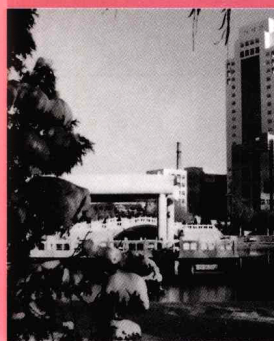




高校社科文库
University Social Science Series

教育部高等学校
社会科学发展研究中心



《红楼梦》中英文语料库 的创建及应用研究

刘泽权/著

The Construction and Applied Studies of
the Chinese-English Parallel Corpus of
Hong Lou Meng

光明日报出版社



高校社科文库
University Social Science Series

教育部高等学校
社会发展研究中心

汇集高校哲学社会科学优秀原创学术成果
搭建高校哲学社会科学学术著作出版平台
探索高校哲学社会科学专著出版的新模式
扩大高校哲学社会科学科研成果影响力



刘泽权/著

《红楼梦》中英文语料库 的创建及应用研究

The Construction and Applied Studies of
the Chinese-English Parallel Corpus of
Hong Lou Meng

光明日报出版社

图书在版编目(CIP)数据

《红楼梦》中英文语料库的创建及应用研究/刘泽权著.
—北京市:光明日报出版社,2010.4
(高校社科文库)
ISBN 978-7-5112-0696-1
I. 红… II. 刘… III. 红楼梦—英语—翻译—研究
IV. H315.9

中国版本图书馆 CIP 数据核字(2010)第 057889 号

《红楼梦》中英文语料库的创建及应用研究

作 者: 刘泽权 著

出 版 人: 朱 庆

责任编辑: 田 苗

封面设计: 小宝工作室

责任校对: 田 璐 施凌波 刘家勇

责任印制: 胡 骑 宋云鹏

出版发行: 光明日报出版社

地 址: 北京市崇文区珠市口东大街 5 号,100062

电 话: 010-67078945(发行),67078243(总编室),67078235(邮购)

传 真: 010-67078227,67078233(发行),67078255(办公室)

网 址: <http://book.gmw.cn>

E - mail: gmcbs@gmw.cn

法律顾问: 北京市华沛德律师事务所张永福律师

印 刷: 北京大运河印刷有限责任公司

装 订: 北京大运河印刷有限责任公司

本书如有破损、缺页、装订错误,请与本社发行部联系调换

开本: 690×975 毫米 1/16

字数: 275 千字

印张: 15.5

版次: 2010 年 5 月第 1 版

印次: 2010 年 5 月第 1 次印刷

书号: ISBN 978-7-5112-0696-1

定价: 35.00 元

版权所有 翻印必究



前言

“《红楼梦》中英文语料库的创建及应用研究”系2005年度国家社会科学基金项目（项目批准号：05BYY028），资助金额7万元，研究历时两年半（2005年5月至2007年12月）。该语料库包括蔡义江校注、浙江文艺出版社1993年出版的一百二十回《红楼梦》，1892至1893年在香港出版、由前英国驻澳门副领事乔利（Bencraft Joly）翻译的二卷本 *Hung Lou Meng; or, The Dream of the Red Chamber*（前五十六回），由英国汉学家霍克思（David Hawkes）及其婿闵福德（John Minford）翻译、英国企鹅出版社自1973至1986年间出版的一百二十回 *The Story of the Stone*，以及由中国翻译家杨宪益与他的英国夫人戴乃迭（Gladys Yang）合译、中国外文出版社于1978至1980年出版的一百二十回 *A Dream of Red Mansions*。整个项目的完成经过了语料的收集、加工、标注、对齐、检索统计及初步应用等阶段，现已顺利结题。回顾项目进行期间的工作和努力，我们感受最深的大概可以凝结成两点——艰辛却富有意义。

一方面，随着科学技术的不断发展，计算机的广泛应用，通过建立语料库进行语言研究的方法得到了人们的普遍认可。“《红楼梦》中英文平行语料库”的建成有利于研究者对《红楼梦》的中英文本进行语内及语际对照研究，方便国内外学者充分利用语料库储存数据大、计算机运行速度快、语料库工具能准确对大量数据进行分析的诸多优势，系统、全面地对三个英译本在翻译技巧运用、文体设计、语言选择与文化内容处理等方面进行对比分析，从而总结归纳类似《红楼梦》的汉语典籍英译的一般规律、译者风格，检验翻译普遍性的表现，进而指导我国的典籍翻译理论研究和教学实践。同时，该语料库的出版发行与使用，对于全面研究《红楼梦》的语言运用，尤其是对比前八十回与后四十回在文体上的种种差异，系统检索全书的诗词曲赋、成语典故、修辞技巧、官职称谓、宗教礼仪、人名地名、饮食服饰及其不同的英文表达，编纂



相关的一文多译的《红楼梦》汉英词典,传播中华典籍,弘扬中华文化等方面都会起到积极的推动作用。此外,中外五位译者呕心沥血所成就的三种《红楼梦》英译本,可以作为平行语料库使用,以便于从历时与共时、定性与定量的角度来研究一个多世纪以来的英语语言变化和翻译技巧的变迁,从而更全面和准确地描述各个译本的长短优劣与异同得失之处,进而为翻译评价提供一个可检验的标准,也为汉外翻译教学与实践提供借鉴。因此,我们相信我们所做的工作和付出的努力是很有意义的。另外,为了使“《红楼梦》中英文平行语料库”更加完善,包含的译本更为全面,我们又于近期对《红楼梦》另一个鲜为人知、研究较少的一百二十回全译本进行了语料转化、校对、标注与对齐。该译本由英国传教士、伦敦大学文学博士邦索尔(Bramwell Bonsall)翻译完成,可惜因为种种原因,未能出版。目前,对该译本的处理已经基本完成,很快将充实到已建成的语料库中。

然而,“《红楼梦》中英文平行语料库”涉及汉英两种语言、三(四)个英译本,因此在其建设过程中遇到了诸如底本选择、版本比对、句子划分、属性标注、程序开发、软件应用等实际问题。课题组在这些方面都做了一些积极探索和有益尝试。我们将建库的过程及对所建成语料库的初步应用整理成册,谨期为《红楼梦》英译的相关研究及我国其他典籍的翻译研究提供借鉴和帮助。本书分为两部分:上篇主要围绕“《红楼梦》中英文平行语料库”的创建,着重介绍建库过程、语料库检索使用、数据统计及其它相关问题;下篇为基于语料库的一些初步应用研究,其中大部分已经或即将以独立的论文形式在国内外学术刊物发表。部分章节的内容在国际国内会议上进行了宣读交流,并根据与会专家学者的意见进行了修改。

在项目进展期间,课题组得到了许多国内外专家学者的指导和帮助。特别是在2008年年初为项目举行的成果论证会上,来自我国文学、翻译及语料库语言学领域的各位专家,如中国社会科学院的顾曰国教授、北京外国语大学的王克非教授、南开大学的王宏印教授、北京大学的王逢鑫教授及中国翻译协会会刊《中国翻译》副主编杨平教授,听取了课题组对项目进展和完成情况的汇报。在对项目充分肯定的基础上,专家组还对其所存在的问题及进一步改进和发展的方向提出了中肯的意见和建议。另外,国家教育部语言文字应用研究所的冯志伟研究员、英国伯明翰大学教授暨《国际语料库语言学》主编 Wolfgang Teubert、上海交通大学的卫乃兴教授、大连外国语学院徐璐教授以及北京外国语大学的许家金博士也都先后,甚至多次到访课题组讲学指导,与课



题组成员进行了广泛深入的交流与探讨,使我们受益匪浅。在此,我们对先后莅临课题组的各位专家表示深深的感谢。

本项目的完成离不开燕山大学外国语学院多名研究生的辛勤工作和无私付出。他们是2004级硕士研究生田璐、古月、侯羽、李力超、车晓睿、廉张军、白塔娜、邕丽涛,2005级硕士研究生朱虹、王璐、白杨、张瑜、刘超朋、闫继苗、陈银春、苗海燕、范哲巍,2006级硕士研究生张丽和2007级硕士研究生刘鼎甲。以上这些同学在学习、工作和研究之余,牺牲自己的休息时间,包括两个暑假,任劳任怨且热情饱满地投入到课题组的工作当中,从而保证了项目的顺利完成。尤其感谢以下相关章节的主研人员:

第二、七章	刘超朋	第三章	张丽、谭晓平
第四、九章	田璐	第五章	田璐、刘超朋、刘鼎甲
第六章	刘鼎甲	第八、十四章	朱虹
第十章	闫继苗	第十一章	陈银春
第十二章	侯羽	第十三章	苗海燕

燕山大学信息学院博士生导师刘国华教授及其带领的数据库研究小组成员苑迎、刘春辉、麻会东、张宇雪中送炭,对数据库的创建及网络检索平台的开发提供了无私的技术支持。在此向他们一并表示衷心的感谢。

在课题的进展过程中,燕山大学科学技术处、外国语学院及图书馆的领导及相关负责人给予了大量的帮助,燕山大学学校领导也给予了大力的鼓励和支持。燕山大学科技处和外国语学院共同给予本课题五万元的配套资助,购买了计算机、扫描仪、电脑桌椅等必要的设备,建立了当时我校文科唯一的实验室,从而有力地确保了本课题的硬件和平台基础。在此,我向学校和上述相关部门、院系的领导表示诚挚的感谢和崇高的敬意。

“《红楼梦》中英文语料库的创建及应用研究”项目时间紧、任务重。尽管课题组克服了诸多困难,进行了不断的创新与尝试,然而所做的工作还有很多不足之处。我们诚恳接受大家的批评和建议,以使建成的语料库更好地满足大家的科研和教学需要。希望“《红楼梦》中英文平行语料库”能帮助更多的学者利用语料库的统计结果和检索工具,结合大量标注好的语料,得出更有价值、有意义的研究成果。

刘泽权

2009年1月于秦皇岛



CONTENTS 目录

上篇 《红楼梦》中英文平行语料库的创建

第一章 绪论 / 1

第二章 语料库与平行语料库概述 / 4

一、语料库的定义 / 4

二、语料库的发展 / 5

三、语料库的类型 / 5

四、语料库的应用 / 7

五、平行语料库 / 8

第三章 《红楼梦》版本述要 / 11

一、主要版本 / 11

二、回目差异 / 13

三、问题章回 / 15

第四章 《红楼梦》英译述要 / 22

一、主要节译本 / 22

二、全译本 / 24



三、译本研究综述 / 27

第五章 《红楼梦》中英文平行语料库的创建 / 32

一、创建步骤 / 33

二、创建方法 / 33

三、网络检索平台 / 49

四、结语 / 54

第六章 检索软件及应用介绍 / 55

一、基于 WordSmith 的单语纯文本语料检索 / 55

二、基于 AntConc 的单语纯文本语料检索 / 66

三、基于 EditPlus 的语料检索 / 70

四、基于 ParaConc 的中英文平行语料检索 / 73

第七章 《红楼梦》中英文平行语料库数据统计与分析 / 77

一、语料库与译者风格 / 77

二、译本数据统计与分析 / 79

三、结语 / 83

下篇 《红楼梦》中英文平行语料库的应用研究

第八章 《红楼梦》中的习语及其翻译研究 / 88

一、汉英习语的理论阐释 / 89

二、《红楼梦》习语及其翻译研究综述 / 90

三、语料与研究方法 / 91

四、分析讨论 / 92

五、结语 / 100

第九章 《红楼梦》叙事标记语及其英译——基于语料库的对比分析 / 101

一、理论依据 / 101



二、《红楼梦》叙事标记语及其分类 / 103

三、语料分析 / 104

四、结语 / 108

第十章 译者风格与翻译策略——以《红楼梦》报道动词的 英译为为例 / 110

一、报道动词研究述评 / 110

二、语料与研究方法 / 112

三、分析讨论 / 115

四、结语 / 119

第十一章 英译被动式的信息表达功能——以《红楼梦》第一回的四种 英译为为例 / 121

一、被动句式 / 121

二、文献综述 / 122

三、主位结构语篇分析 / 123

四、结语 / 130

第十二章 《红楼梦》三个英译本“死亡”委婉语比较 / 132

一、Leech 的语义划分理论 / 132

二、“死亡”委婉语的分类 / 132

三、《红楼梦》中“死亡”委婉语的英译比较 / 133

四、结语 / 136

第十三章 《红楼梦》“尚红”——基于语料库的语义分析 / 137

一、理论背景 / 137

二、《红楼梦》中的“红” / 138

三、《红楼梦》“尚红” / 140

四、结语 / 147



第十四章 英译刘姥姥形象的对比研究——以话语的人际功能分析

为例 / 148

一、文献综述 / 149

二、语料与研究方法 / 150

三、人际变量阐释 / 151

四、分析讨论 / 152

五、结语 / 164

第十五章 尊他敬语与中华文化——由《红楼梦》的英译谈起 / 166

一、尊他敬语与中华文化综述 / 166

二、语料分析 / 169

三、结语 / 179

第十六章 “大爷”、“嫂子”及其五种英译的人际功能比较 / 181

一、理论建构 / 182

二、语料分析 / 185

三、结语 / 190

第十七章 Translating Tenor: With Reference to the English Versions of *Hong Lou Meng* / 192

I. Introduction / 192

II. Titles / 195

III. Honorifics / 208

IV. Conclusion / 220

《红楼梦》英译书目 / 222

参考文献 / 223



上 篇

《红楼梦》中英文平行语料库的创建

第一章

绪 论

《红楼梦》作为中华文化的“百科全书”，在我国文学领域享有崇高的地位，对于研究我国的文学艺术、历史文化都具有极高的价值，于是红学在国内外蓬勃发展起来。相形之下，对《红楼梦》译本的研究就略显不足，其研究的系统、规模与深度都远远不够理想。就《红楼梦》的英译而言，从1830年至今，出现了11种英译本，“成为中国文学英译史上一道独特而灿烂的风景”（陈宏薇、江帆，2003：46），其中具有代表性的几种译本是：英国人乔利（Joly, 1892~1893）的前56回（全）译本（以下简称乔译）；英国汉学家霍克思和闵福德（Hawkes & Minford, 1973~1986）的120回全译本（以下简称霍译）；中国翻译家杨宪益及其夫人戴乃迭（Yang & Yang, 1978~1980）的120回全译本（以下简称杨译）以及时任哥伦比亚大学中文教授的王际真（Wang, 1929; 1958）、英国的麦克休姐妹（McHugh & McHugh, 1958）等翻译的若干节译本。^①

^① 《红楼梦》英译本情况详见第四章。



由于《红楼梦》早期的英译本很不完整,仅是片断翻译,所以对它们的研究很少。目前绝大部分的红译研究以霍译和杨译为对象的多,对多个译本进行全面系统研究的少;基于人工分析的研究多,基于语料库的研究还较缺乏。这种局面的出现,不仅使其他一些同样很有价值的译文备受冷落,而且导致研究所得出的结论难免失之偏颇(Liu, 2006: 55),在一定程度上缺乏可靠性,也不具推广性。当然,这些问题的存在,除了研究者的主观因素外,更主要的原因恐怕还在于:(1)由于资源共享的不畅通,目前大多数研究者只能找到有限的一两个译本;(2)《红楼梦》篇幅长、内容广、译文多,单个的研究者很难有时间、精力和财力进行大规模且具有深度和广度的研究;(3)现代科学技术支持和运用的缺乏。

随着科学技术的不断发展,计算机的广泛应用,通过建立语料库进行语言研究的方法得到了人们的普遍认可。基于语料库的研究主要具有三大优点:“(1)运用计算机的强大功能,进行快速、准确和复杂的分析;(2)语料库规模大、语域宽、语言信息范围广;(3)既有定量分析,又有定性解释功能,对语言的描写比较全面。”(王克非等, 2004: 4~5)因此,语料库已被广泛应用到信息查询、语言教学、词典编纂、翻译研究等各个领域,涉及语言的词汇、语法、篇章、文化等不同层面的研究。

平行语料库作为语料库的一种,是指“由原文文本及其平行对应的译语文本构成的双语语料库”(王克非, 2004: 4~6)。平行语料库让人们认识到“翻译研究应从规约性(prescriptive)研究向描述性(descriptive)研究过渡”(Baker, 转引自廖七一, 2000: 380),从而使得翻译研究更具有普遍性和客观性,研究方法和得出的结论也更加科学可靠。

建立一个科学可靠的《红楼梦》中英文平行语料库,能够方便翻译学者充分利用语料库储存数据大、计算机运行速度快、语料库工具能准确对大量数据进行分析的优势,系统、全面地对《红楼梦》各译本在翻译技巧运用、文体设计、语言选择与文化内容处理等方面进行对比分析。通过定量分析和功能阐释相结合,获得各种客观实用的研究成果(Biberetal, 1999: 233),指导我国的翻译实践、翻译教学与其他跨文化交际研究。

建立和开展以语料库为基础的翻译研究一定要严格遵循基于语料库的翻译研究方法并经过以下步骤。首先是文本收集,也就是收集现有的《红楼梦》原文及其译本,通过扫描、校对,变成可机读的文本。其次是双语语料的预处理,即对收集的双语文本进行必要的预处理,主要包括文本净化、中文部分的



切分及英文部分的断词等。第三，建立双语平行语料库。中文文本和英文文本在句子一级进行对齐，通过段与段、句与句的对齐把中英文语料系统地编成可索引的数据。第四，建立等价单位的索引。为了获取汉英翻译等价单位，在句与句对齐的基础上，对中英文语料进行翻译等价单位（TEP）的标注。一个翻译等价单位是一个二元组（STU，TTU），其中 STU 是一个源语言翻译单位，TTU 是一个目标语言翻译单位，STU 和 TTU 之间具有翻译等价关系，也就是说 STU 以及 TTU 之间具有互译关系。第五，实现等价单位的自动抽取。为了能够抽取研究者感兴趣的中英翻译等价单位，需要建立自动检索工具，这些工具包括可单独运行的检索软件和基于网页的检索。

建立“《红楼梦》中英文平行语料库”，有利于《红楼梦》译学的发展，同时也与许多红学家所倡导的通过“红楼梦网络资料中心”来促进红学发展的设想相得益彰（罗凤珠，1997）。该语料库的使用与出版发行，对传播与弘扬中华民族的文化传统，使《红楼梦》这一中华文库的瑰宝在异域大放异彩，也会起到积极的推动作用。另外，在“《红楼梦》中英文平行语料库”建设与应用研究方面所获得的成就与经验，将会为《红楼梦》其他语种和其他典籍的翻译研究提供借鉴和帮助。



第二章

语料库与平行语料库概述

语料库语言学 (Corpus Linguistics) “通过语料库 (现在一般是电子语料库), 对语言进行研究” (Chalker & Weiner, 1994: 98)。语料库语言学在过去的几十年中得到了长足的发展, 现在已成为语言学的一个重要分支。然而对于语料库的界定和分类, 语料库对语言研究所起的作用以及是否应该把语料库语言学列为一个独立的学科, 学术界仍存在不同程度的分歧。

一、语料库的定义

“语料库来自拉丁词 ‘corpus’, 原意为 ‘汇总’、‘文集’ 等” (杨惠中, 2002: 35)。“语料库 (corpus)” 其英语单词的复数形式为 “corpora” 或者 “corpuses”。在语言学领域, 语料库并不能简单指 “文集” 或 “汇总”, 这一点在很多语言学家中取得了共识, 但具体的学者还是对其定义有着细微的不同。

Baker (1995: 226) 认为语料库一般用来指 “电子形式的可以进行自动或者半自动分析的文本”。而 Tognini - Bonelli (2001: 2) 的观点是: “语料库可以定义为用于语言学分析的具有代表性的文本的集合……它的研究对象为自然的、真实的语言。” Biber 等 (1998: 246) 则表述为 “一个语料库不只是一些文本的集合, 它是用来试图代表一种语言或者一种语言的一部分。语料库的合理设计将取决于它试图体现的语言特点。”

对于语料库的基本属性, McEnery 和 Wilson (1996: 21) 列出了如下几点: 代表性、确定的规模、可以机器阅读的形式、可以作为一个标准的参考。俞士汶 (2003: 76) 则认为 “在现代语言学的意义上, 对语料库的认识至少应该包含如下三点: (1) 语料库中存放的是在语言的实际使用中真实出现过的语言材料; (2) 语料库是以电子计算机为载体承载语言知识的基础资源; (3) 真实语料需要经过分析、加工、处理, 才能成为有用的基础资源。”

尽管上面不同的学者有各自不同的定义, 并且存在细微的差异, 但我们可



以归纳出他们共通之处，即语料库从现代意义讲是用电子文本存储的用于语言研究的真实、自然语言的集合。

二、语料库的发展

从1959年到1961年 Quirk 进行了他的英语用法调查语料库 (Survey of English Usage Corpus)，广泛收集了当时大众使用的英语语篇材料（包括口头的）。后来陆续出版的两部语法巨著《当代英语语法》（1972）和《英语语法大全》（1985）便是基于此语料库编纂而成。Leech 和其所在的兰卡斯特大学（Lancaster University）从20世纪70年代开始一直是全球语料库语言学的研究基地，他们在1978年与挪威的奥斯陆大学（Oslo University）和勃艮大学（Bergen University）联合创建了世界上第一个百万词次的英语语料库 LOB。1987年与英国 IBM 公司联合建成 Lancaster 大学计算机语料库中心并开始研制词性和句法标注的自动附码系统；1994年又与英国图书馆、Oxford 大学、Longman 出版公司等联合建成了上亿词次的英国国家语料库 BNC。这个语料库“包括2亿多英文单词并且数量一直都在增加”（Leech, 1991: 11）。“正是由于他们长期不懈的、富于创造性的努力，才引导和促进了近30年来语料库语言学异常迅猛的发展”（何安平，2001: F23）。

根据 Leech 的观点，语料库的发展分为三代。第一代是以 Brown 语料库为代表，该语料库拥有上百万词，在当时看来这样的规模已经是相当巨大了。20世纪80年代的第二代语料库是以 John Sinclair 的伯明翰大学英语文本集和朗文 - 兰卡斯特英语语料库为代表。Leech 认为第三代语料库应该指那些拥有上亿词次的语料库。按照 Leech 的标准，第三代语料库可能就指的是上面提到的英国国家语料库 BNC 了。20世纪90年代以来语料库语言学快速发展，其趋势主要表现在以下三个方面：“大规模、多品种语料库的建设；对语料库的深加工研究；语料库广泛应用于与语言相关的各个领域”（同上：F23）。

我们可以看出，语料库的建设从规模上已经有了巨大的突破，大型语料库的建设已经成为了现实。然而，怎样对语料库进行加工和深层次开发或许是更加复杂和艰巨的任务。不同类型的语料库特点和用途也各有千秋，接下来就一起看一下语料库的具体分类。

三、语料库的类型

根据不同的标准，语料库可以分为不同的类型。下面列举几种比较常见的分类方法，并对一些有争议的概念进行简单的比较分析。



1. 通用语料库和专用语料库

这种分类方法是从研究的目的出发的。通用语料库 (general corpus) 指“收集的语料只是建立一个用于研究的平台, 而并不是针对某个具体的目的”, 而专用语料库则是“设计用来做专门用途研究的”(Kennedy, 1998: 19 ~ 21)。上面提到的 Brown 语料库、LOB 语料库以及英国国家语料库 BNC 都是属于通用语料库, 因为它们收集语料并不是用于某个专门语言研究用途, 而基于这些语料库可以进行多方面的语言特征研究。Leech (1992: 112) 把训练语料库 (training corpus) 和测试语料库 (test corpus) 列为专用语料库之内并且对这两种语料库的发展进行了描述。杨惠中教授等的“上海交通大学科技英语语料库”也属于专用语料库的范畴。

2. 生语料库和标注语料库

这种分法是根据语料库中的语料是否经过处理和标注而决定。生语料库 (raw corpus) 是指没有经过任何加工处理的纯文本语料库。标注语料库, 顾名思义就是对语料进行了标注。当然标注也是分很多的层次。杨惠中列举了几种层次: (1) 经过格式属性标注, 如段落、字体、字号进行标注; (2) 对识别信息进行标注, 如作者、体裁、语域以及词性等标注; (3) 特殊标注, 如错误赋码等。经过标注了的语料库固然有其优越之处, 如可以有针对性地对某项参数进行研究, 但是“不经任何人工介入的生语料库同样具有独特的价值”(杨惠中, 2002: 43)。

3. 单语语料库、双语语料库和多语语料库

这三种语料库是根据语料库中使用语言的种类区分的。这几种语料库从字面上就可以得出各自的特点。值得说明的一点是双语语料库和多语语料库又分为平行语料库 (parallel corpus)、类比语料库 (comparable corpora) 和翻译语料库 (translational corpus)。平行语料库是“由原文本及其平行对应的译语文本构成的双语语料库, 其双语对应程度可有词级、句级和段级几种”(王克非等, 2004: 6)。本研究所建的语料库就属于句子层面对应的汉英双语语料库。类比语料库是“由不同语言的文本或同一种语言不同变体的文本所构成的两个或两个以上的语料库”(同上: 7)。从定义上我们可以看出, 类比语料库侧重于语料库中语料的关系, 从而也有单语类比语料库。翻译语料库是“由彼此具有翻译关系的不同语言文本构成的语料库”(同上)。

4. 其他类型的语料库

除了上面的几种分类方法外, 根据语料的语域特征可以分为书面语语料库



和口语语料库,根据语料状态可分为共时语料库和历时语料库等。语料库的这些分类是从不同的角度提出的,所以很有可能某一个语料库会具有几个不同的属性。此外,不同的理论学家对某些分类的定义存在分歧。例如 Granger (1996: 38) 把运用同样的取样框架建成的几个小语料库称作平行语料库,而按照上文中所提到的标准,这种类型的语料库应该被称为类比语料库。不同的定义各有其合理性,比如 Granger 的定义是从语料库的整体而言,处于平等的位置,而上文的定义则是从语料库的内容之间相互关系来看的。所以在进行语料库语言学研究时要注意前后所采用的分类定义的一致性。

四、语料库的应用

语料库语言学重点在于“语言运用而不是语言能力,是通过观察语言现象得出结论和规律而不是相反”(Leech, 1992: 107)。基于语料库的研究主要具有三大优点:(1)运用计算机的强大功能,进行快速、准确和复杂的分析;(2)语料库规模大、语域宽、语言信息范围广;(3)既有定量分析,又有定性解释功能,对语言的描写比较全面(王克非等,2004: 4~5)。

何安平(2001: F17)列出了以语料库为基础的研究方所涉及的研究领域,如语言结构的核心领域(如词汇学、语法学等),语言学的其他领域(如社会语言学、方言和话语研究、语言教学、文体学、历史文本的比较、语言模式的变异研究等)。王克非等(2004: 5)指出:“作为一个新兴研究领域,一种新的探寻语言本质的哲学方法,80年代以来,语料库语言学发展迅速,对语料库的规模、类型、功能等也提出了更高的要求。……随着语料库语言学的发展,词汇学、语言教学、翻译、文体学、语法学、计算语言学等研究领域将从中受益。”

语料库一个很重要的应用就是词典编纂。广受好评的 COBUILD 英语词典便是基于其英语语料库(Bank of English)编纂而成的。词典中提供大量的例证,取材广泛,都是直接取自 COBUILD 英语语料库,从而使例证真实、地道和新颖,这是其区别于传统词典的独特之处。由于现代语料库是以电子文本存储的,可以被机器识别,所以能极大提高运用语料库进行分析的速度和效率。“语言学的所有领域几乎都可以从语言使用的角度进行研究,因此都是语料库语言学的用武之地”(何安平,2001: F17)。

对于语料库语言学是否应该被称作一个独立的学科,不同语言学家有不同的观点。Tognini - Bonelli 对语料库语言学的性质进行了重新审视,指出语料库语言学并不是一个真正意义上的科学研究领域,只不过是给语言研究提供了一