

严
明
著

商务英语语料库 开发与应用研究

THE STUDY OF
BUSINESS ENGLISH CORPUS:
DEVELOPMENT AND APPLICATION

BUSINESS
ENGLISH



中国商务出版社
CHINA COMMERCE AND TRADE PRESS

商务英语语料库 开发与应用研究

严 明 著

中国商务出版社

图书在版编目 (CIP) 数据

商务英语语料库开发与应用研究/严明著. —北京
中国商务出版社, 2011. 3

ISBN 978 - 7 - 5103 - 0427 - 9

I. ①商… II. ①严… III. ①商务—英语—语言学—
研究 IV. ①H31

中国版本图书馆 CIP 数据核字 (2011) 第 028534 号

商务英语语料库开发与应用研究

Shangwu Yingyu Yuliaoku Kaifa Yu Yingyong Yanjiu

严 明 著

出 版: 中国商务出版社

发 行: 北京中商图出版物发行有限责任公司

社 址: 北京市东城区安定门外大街东后巷 28 号

邮 编: 100710

电 话: 010—64269744 (编辑室)

010—64266119 (发行部)

010—64263201 (零售、邮购)

网 址: www.cctpress.com

邮 箱: cctp@cctpress.com

照 排: 中国农业出版社印刷厂

印 刷: 北京密兴印刷有限公司

开 本: 850 毫米×1168 毫米 1/16

印 张: 24.5 字 数: 549 千字

版 次: 2011 年 3 月第 1 版 2011 年 3 月第 1 次印刷

书 号: ISBN 978 - 7 - 5103 - 0427 - 9

定 价: 40.00 元

版权专有 侵权必究

举报电话: (010) 64242964

前 言

近年来,语料库语言学作为新兴学科得到了蓬勃发展。语料库是为一个或多个应用目标而专门收集的,有一定结构的、有代表性的、可被计算机程序检索的、具有一定规模的语料的集合(冯志伟,2009)。语料库的应用涉及多个领域,如语言教学、语言分析、词典编纂和测试设计等。这种基于大量真实语料的语言研究方法通过研究系统选取的自然语言样本,揭示语言使用规律,尤其对于商务英语教学模式与方法颇有启发。

黑龙江大学外语教学研究部始于1985年,并于2007年成立应用外语学院。近年来,通过全体院部同仁的团结奋进和努力拼搏,黑龙江大学成为国家首批大学英语教学改革示范点项目单位。2008年,教育部正式批准黑龙江大学开设商务英语本科专业。随后,我们先后取得外国语言学及应用语言学硕士和翻译专业硕士(MTI)两个硕士学位授予权。2009年,我们的团队获得国家级教学成果二等奖。

2007年,黑龙江大学应用外语学院与外语教研部开始了基于语料库的教学模式探索与研究。2007年年底,我们正式启动黑龙江大学商务英语语料库(Heilongjiang University Business English Corpus,以下简称HUBEC)的建设项目。经过认真细致的调研,并参考国内外同类研究的经验,我们最终确定了HUBEC的总体建设思路:5年之内建成可以用于商务英语本体研究、商务英语教学研究、商务英语词典编纂和商务英语测试的大型商务英语笔语语料库,库容5 000万词次。基于上述理念,HUBEC的总体目标如下:

第一阶段(2009年7月至2010年7月):库容量1 000万词次;

第二阶段(2010年9月至2011年5月):库容量3 000万词次;

第三阶段(2011年7月至2012年7月):库容量5 000万词次。

2010年年底,HUBEC的规模已经突破1 500万词次,基本具备

了可持续发展的能力。在此基础上,我们提炼并总结出《商务英语专业(本科)参考词汇表》、《商务英语专业(本科)基础参考词汇表》、《商务英语专业(本科)术语参考词汇表》和《商务英语专业(本科)缩写词表》。它们是 HUBEC 在商务英语研究方面所做的初步探索。这些词汇表将有助于我们提高对商务话语本体的认识,为商务英语人才语言能力构成及培养模式提供有力的数据支撑。同时,我们也非常渴望与国内外的同仁积极交流,共同为国内商务英语专业的发展贡献力量!

本书共有七章:第一章语料库语言学概述,主要回顾了语料库语言学发展历程、理论依据、研究本质与特点,以及国内外语料库语言学研究现状;第二章主要介绍了语料库的一些基本概念;第三章是关于语料库建设的基本流程与关键问题;第四章主要讨论了基于语料库的商务英语相关理论;第五章总结了语料库的主要应用领域;第六章详细介绍了黑龙江大学商务英语语料库的建设过程;第七章内容为基于 HUBEC 的商务英语词汇表的开发与研究。

本书的出版得到了中国商务出版社的大力支持,出版社编辑人员为书稿的编辑、校对工作也付出了辛勤的劳动,在此我们表示衷心的感谢。

最后,我要感谢我们黑龙江大学应用外语学院年轻的“商务英语语料库开发与应用研究”团队,没有他们不懈的努力和废寝忘食的工作,本语料库及其相关成果是不可能问世的。

限于水平的制约,本书难免存在疏漏和不足。真诚地希望各位专家和学者对本书提出宝贵的批评和建议。

严 明

2010 年 12 月

于黑龙江大学子衿园

本书得到黑龙江省教育厅人文社科面上项目和黑龙江大学
杰出青年科学基金项目“商务英语语料库的设计、开发与应
用”（项目编号分别为：11542167、JC2007W4）的资助。

目 录

第一章 语料库语言学的概述	1
1.1 语料库语言学的历史发展	1
1.1.1 语料库语言学的开端	1
1.1.2 语料库语言学的沉寂	1
1.1.3 语料库语言学的复苏	2
1.2 语料库语言学的理论依据	2
1.2.1 源于经验主义的理论依据	3
1.2.2 源于结构主义的理论依据	5
1.3 语料库语言学的本质	5
1.4 语料库语言学的研究特点	6
1.4.1 语言研究的特点	6
1.4.2 语料库语言学的研究特点	6
1.5 语料库语言学的研究领域	7
1.5.1 语言研究	7
1.5.2 外语教学	8
1.5.3 词典编纂	8
1.5.4 机器翻译	8
1.6 语料库语言学的国外研究现状	8
1.6.1 语言研究	9
1.6.2 语言教学	10
1.6.3 翻译研究	11
1.7 语料库语言学的国内研究现状	12
1.7.1 语言研究	12
1.7.2 翻译研究	13
1.7.3 外语教学	13
第二章 语料库的概述	15
2.1 语料库的定义	15
2.2 语料库的分类	16

2.3 语料库的发展历史	18
2.3.1 国外语料库研究	18
2.3.2 国内语料库研究	20
第三章 语料库的建设	23
3.1 语料库的建设	23
3.1.1 语料库的规划	23
3.1.2 语料库的开发	26
3.1.3 语料的采集	26
3.1.4 语料的抽样	26
3.2 语料库建设中的核心问题	27
3.2.1 语料库的代表性	27
3.2.2 语料库的平衡性	28
3.2.3 语料库的时间性	28
第四章 基于语料库的商务英语理论研究	29
4.1 商务英语与专门用途英语	29
4.1.1 专门用途英语的定义	30
4.1.2 商务英语在专门用途英语体系中的定位	32
4.2 国际商务相关概念	35
4.2.1 商务与 business	35
4.2.2 国际商务与 international business	36
4.2.3 国际商务领域的界定	38
4.3 国际商务英语的概念与研究角度	41
4.4 基于体裁的商务英语语料库中的语料分类研究	41
4.4.1 体裁的定义	41
4.4.2 体裁的分类研究	42
4.5 基于语料库的测试理论研究	44
4.5.1 语言测试的实质和根本要求	44
4.5.2 特殊目的语言测试	45
4.5.3 商务英语测试	47
4.5.4 商务英语测试语料的收集	48
第五章 语料库的应用	49
5.1 语料库与语言研究	49

5.1.1 语料库与词汇	49
5.1.2 语料库与语法	50
5.1.3 语料库与话语分析	51
5.1.4 语料库与语义学	52
5.1.5 语料库与社会语言学	52
5.1.6 语料库与文体学	53
5.1.7 语料库与历史语言学	53
5.2 语料库与语言教学	54
5.2.1 教学大纲设计	54
5.2.2 英语教材编写	54
5.2.3 教学模式的改革	55
5.2.4 课程教学	56
5.3 语料库与词典编纂	57
5.4 语料库与测试	58
5.5 语料库与翻译	59
5.5.1 语料库与人工翻译	59
5.5.2 语料库与机器翻译	60
第六章 黑龙江大学商务英语语料库的建设与开发	61
6.1 黑龙江大学商务英语语料库的建设目的	61
6.2 HUBEC 的建设思路与目标	64
6.3 商务英语语料库和学术英语语料库概述	65
6.3.1 Cambridge International Corpus (CIC)	65
6.3.2 PolyU Business Corpora 和 Hong Kong Financial Services Corpus (HKFSC)	66
6.3.3 Business Letter Corpus (BLC)	66
6.3.4 Business English Corpus (BEC)	66
6.3.5 Wolverhampton Business English Corpus (WBEC)	68
6.3.6 The Academic Corpus (AC)	69
6.3.7 The Reading Academic Text corpus (RAT)	69
6.3.8 Michigan Corpus of Academic Spoken English (MICASE)	70
6.3.9 The Corpus of Professional English (COPE)	70
6.4 HUBEC 建设的理论综述	70
6.4.1 理论依据	71
6.4.2 商务英语学科 (Disciplinary) 类书面语语料库构成	72
6.4.3 商务英语实务 (Professional) 类书面语语料库构成	73

6.4.4 商务英语语言教学类书面语语料库构成	74
6.4.5 商务英语学科与语言教学类术语语料库构成	74
6.5 HUBEC 的整体结构	74
6.5.1 HUBEC 的性质	74
6.5.2 HUBEC 的整体结构及规模	74
6.6 HUBEC 的开发意义	76
第七章 商务英语专业词汇表的开发与研究	77
7.1 国内外英语词汇表研究综述	78
7.1.1 通用英语词汇表 (General Service List, GSL)	78
7.1.2 大学英语词汇表 (University Word List, UWL)	78
7.1.3 学术英语词汇表 (Academic Word List, AWL)	79
7.1.4 牛津常用 3 000 词和朗文基础 3 000 词	81
7.1.5 纳尔逊商务英语词汇表	81
7.1.6 国内外英语词汇表相关研究	83
7.2 基于 HUBEC 的商务英语专业参考词汇研究	83
7.2.1 商务英语专业 (本科) 参考词汇表	85
7.2.2 HUBEC 专业术语表	85
7.2.3 商务英语专业 (本科) 参考词汇词频表	86
7.2.4 HUBEC 词频表	86
结束语	87
附录 1: 商务英语专业 (本科) 参考词汇表	88
附录 2: 商务英语专业 (本科) 基础参考词汇表	232
附录 3: 商务英语专业 (本科) 术语参考词汇表	355
附录 4: 商务英语专业 (本科) 缩写词表	360
附录 5: 《商务英语语料库开发与应用研究》光盘使用说明	369
参考文献	371

第一章 语料库语言学的概述

语料库语言学自 20 世纪中叶在欧美国家兴起，至今已走过六十余年的发展历程，作为语言学研究领域中的一种全新方法，它以随机或依照概率方法采集的语言样本为研究对象，以定量与定性相结合的方式分析语言事实，旨在揭示自然语言的规律性特征。

1.1 语料库语言学的历史发展

语料库语言学是数据收集与语言学理论研究的有机结合 (Halliday, 2001)，是语言学理论与应用研究的一场巨大变革。近年来，随着语料库设计、开发和建设的配套技术飞速发展，语言学者对语料库语言学的关注和研究进入了一个新的高潮。当然，语料库的发展并非一帆风顺，也曾经遭受过质疑和打击。简单地说，它的发展历程大致经历了以下几个阶段。

1.1.1 语料库语言学的开端

根据 T. 麦克内里 (T. McEnery) 和 A. 威尔逊 (A. Wilson) (1996) 的研究，19 世纪末 20 世纪初的语言学家们已经采用了一种类似语料库的研究方法，尽管与语料库研究方法存在本质的区别，但这种方法却是现代语料库语言学研究方式的启蒙和开端。该方法主要应用于语言习得、音位学和方言学等相关领域。在语言习得领域，早期欧洲的语言学家十分热衷儿童语言习得的研究，他们把父母详细记载的子女语言发展情况作为进一步语言学理论研究的主要依据；在音位学领域，语言学家则强调以自然、客观的手段获取和分析语料；在方言学研究中，研究人员早期采集方言材料时所使用的取样方法构成了目前语料采集方式的雏形。

K. 艾杰默 (K. Aijmer) 和 B. 阿尔腾伯格 (B. Altenberg) (1991) 认为，以真实语料为基础的语言研究始于 L. 布卢姆莫尔德 (L. Bloomfield) 之后的结构主义。他们指出，那个时期的语言学家受实证主义和行为主义理念影响，强调以语言中自然出现的数据作为语言研究的出发点，反对完全依靠直觉和心智的例证。

1.1.2 语料库语言学的沉寂

20 世纪 50 年代，在实证主义和行为主义思潮的影响下，经验主义成为语言学研究的主导思想。这种学术导向使基于语料的语言学研究成为当时的焦点。60 年代初，随着 N. 乔姆

斯基 (N. Chomsky) 一系列著作的发表, 转换生成语法逐渐占据了学术界的主导。乔姆斯基 (1965) 认为, 语法理论的关键不在于证据的缺乏, 而是当今的语言理论不足以对大量的证据提供恰当的解释。他还指出, 语言学应当研究人脑的语言机制, 即语言能力, 而不是语言的运用。语言能力存在于理想的本族语者的头脑中, 无法直接观察。能够直接观察的语料, 是语言运用的结果, 不能用来揭示语言的本质。转换生成语法理论反对研究具体语料, 这改变了传统的实证研究方向, 语料库在语言研究中的作用也因此遭到了质疑。此后, 转换生成语言学家采用的内省法 (introspection) 成为主流, 这种方法也被称为基于语言能力的研究方法 (competence-based approach)。它以“语言学家自身为研究素材的提供者 (informants), 依靠自己的语感 (intuitions) 作为判断语言现象歧义、正误和可接受性等的依据” (杨惠中, 2002: 6)。以乔姆斯基为首的转换生成语法学派的各种批评从根本上改变了 20 世纪 60 年代结构主义语言学的研究方向, 在此后的 20 年中, 语料库研究方法遭受了沉重打击, 进入低谷。

在当时的语言学界, 除了乔姆斯基外, D. 阿伯克龙比 (D. Abercrombie) (1963) 也曾指出, 语料库方法是一种错误的研究。这主要因为在当时的技术条件下, 没有成熟的手段支持大量的语料搜集、加工和研究。

1.1.3 语料库语言学的复苏

虽然遭受质疑和批评, 但以 R. 夸克 (R. Quirk) 和 W. 弗朗西斯 (W. Francis) 为代表的学者从未停止探索的脚步。夸克强调自然语言材料的重要性, 认为语料库语言学的研究是对复杂的自然现象的一种简化。他于 1959 年开始建立并完成了“英语用法调查”语言资料库 (Survey of English Usage, 简称 SEU)。该库中大量收集了风格题材各异的英语口语和笔语语料, 为进一步的语言系统描写奠定了基础。在随后的 40 年中, 夸克不断扩大语料库语言学的研究范围, 巩固并提高了语料库研究在语言学界的地位, 他的努力对语言理论的研究也产生了深远的影响。1961 年, 弗朗西斯和他的团队建成了世界上第一个机读语料库——布朗语料库 (Brown Corpus)。上述两个语料库是现代语料库语言学的开端。

1975 年, 以 T. 司威克 (T. Svartvik) 为首的语言学家对夸克语料的口语部分进行了标注, 并最终实现了机读, 建立了伦敦-隆德语料库 (London-Lund Corpus)。此后, 沉寂了近二十年的语料库语言学开始复苏。随着计算机科学的不断发展, 越来越多的语言学者加入到语料库语言学的研究行列, 语料库语言学逐渐成为语言学家验证假设、完善理论的重要方法和途径。

1.2 语料库语言学的理论依据

一般认为, 语料库语言学的兴起与发展受到两种思想的影响: 一个是 20 世纪 60 年代初

英国经验主义语言理念；另一个是美国结构主义语言学传统。

1.2.1 源于经验主义的理论依据

语料库语言学的发展在很大程度上受到英国伦敦语言学派的影响。伦敦语言学派又称弗斯(Firth)语言学,是一种源于早期经验主义研究理念的语言学流派,它的主要代表人物是弗斯。

1.2.1.1 经验主义(Empiricism)

英国著名哲学家B. 拉塞尔(B. Russell)(1976)用倒立的金字塔比喻理性主义(Rationalism)的研究方法,用正立的金字塔比喻经验主义的研究传统,从中可以看出早期理性主义和经验主义的研究模式存在巨大的差异。冯志伟在《应用语言学中的语料库》的导读中指出,“20世纪50年代以前,现代语言学的传统,无论是规范语言学、历史语言学或是描写语言学都注重语言事实,提倡经验主义”(Hunston, 2006: 11)。

文艺复兴后期重视实验和分析方法的欧洲自然科学得到迅猛发展,从而影响了近代经验主义理论的提升。从古希腊的N. 亚里士多德(N. Aristotle)到17世纪的F. 培根(F. Bacon)和J. 洛克(J. Locke),再到18世纪的G. 伯克利(G. Berkeley)和D. 休姆(D. Hume),他们都“强调感觉经验的重要性,认为一切知识都来自感觉,否认天赋观念,倾向于对事物进行孤立的分析研究,强调客观依据和实用价值”(陈勇, 2003: 58—59)。“感觉优先原则”是经验主义论证的核心,它强调研究基础的客观性,认为“只有感觉经验的知识才是科学知识;没有感官经验来源的知识都不是科学知识”(孙志海, 2005: 18)。由此可见,基于经验主义的语言学研究重视源于感觉、经验和与客观世界的实证交互,这种理念推动了对语言的起源和发展等问题的进一步深入剖析和考证,也为以后辛克莱(Sinclair)和M. A. K. 韩礼德(M. A. K. Halliday)等人在词汇和语篇,以及语料库语言学方面的研究提供了哲学理据。与经验主义相对的是产生于19世纪晚期的理性主义,理性主义者认为理性是识别、判断、评估实际理由以及使人的行为符合特定目的等方面的智能(Robins, 1997)。乔姆斯基的转换生成语法实际上是对理性主义哲学思想的继承,他认为统计不能解释语言的内在规则或生成机制,为根本解释自然语言(而非形式语言)需寻找充分必要条件。他还试图发现自然语言作为人脑认知器官的一个重要组成部分的内在机制(周允程, 2008)。转换生成语法主要研究心智内部的语言生成模型,故使用由个人直觉提供的内省数据。在转换生成语法形成的初期,乔姆斯基的观点与传统的经验主义哲学范式大相径庭,更多地体现为理性主义、心灵主义(Mentalism)和自然主义(Naturalism)的思想(周允程, 2008)。他的思想是对传统的经验主义哲学的超越,一经提出就被多数学者所接受。反之,受经验主义影响较大的语料库语言学则受到了质疑。

然而,人们逐渐发现理性主义研究所得出的语言规则不能得到广泛推广,甚至乔姆斯基

也开始改变初衷,注重对具体词汇的研究。显然,基于理性主义和经验主义的研究方法和认识理念都过于极端,不利于语言学的发展。如果将经验主义的定量分析和理性主义的定性分析结合起来,充分利用并整合人类已有的各种经验和资源,将会极大的推动语言学的发展。令人欣慰的是,近现代理性主义和经验主义相互结合的倾向日趋明显,例如 I. 康德(I. Kant)的先验(Priori)理性主义哲学“赋予了经验主义某种合法的地位”,“唤醒了欧洲启蒙思潮所蕴含的科学理性压倒一切、支配一切的理性至上、科学万能倾向的警惕”(张能为, 2010: 31)。此后,更多的学者开始将理性主义的研究理念同实证主义的研究传统相融合,使用按一定标准搜集的自然语料验证语法理论。

由此可见,早期的语料库语言学更多地采用了经验主义的方法论。随着近现代语言学理论的不断发展与融合,语料库语言学的学者们也开始接受理性主义的研究模式,并在实际的语料库建设和基于语料库的语言描述和分析中,不断地渗透理性主义的思维理念。可以说,语料库语言学的产生和发展为经验主义和理性主义的相互支撑、共同发展提供了绝好的契机。此外,二者的进一步结合也会推动语料库语言学研究领域的深化和拓展。

1.2.1.2 弗斯语言学

弗斯是英国现代语言学的奠基人,他的思想影响和熏陶了英国的第一代和第二代语言学家。弗斯的学生继承和发扬了他的语言理论,逐渐形成了伦敦学派,该学派与布龙菲尔德的结构主义语言学以及乔姆斯基的转换生成语法学派共同成为 20 世纪语言学界最具影响力的主流语言学派。

弗斯受培根以来的经验主义认识论的影响较大。他坚持反对心灵主义(Mentalism)的立场,强调使用可观察的证据,主张用实证的方法研究语言,推崇理论和经验的结合,这些都成为弗斯语言学的出发点与核心要素(卫乃兴, 2008)。弗斯的这些思想极大地影响了此后语料库语言学的研究范式——在随机采样的基础上收集有代表性的真实语言材料,然后通过概率的手段对语言的整体规律加以描述和分析。

弗斯的学生韩礼德延续了他的语言学理念,并将之提炼为处理相关语言现象的框架,为系统功能理论奠定了坚实基础。系统功能语言学同语料库语言学一样都强调“自然发生的语言”(naturally occurring language);重视“上下文”(context)。系统功能语言学还提出了“概率”的思想,同时把“频度”作为一个重要的操作性特征。韩礼德本人对语料库语言学也怀有浓厚的兴趣,他在《功能语法导论》(*An Introduction to Functional Grammar*)一书中用一章的内容探讨了基于语料库的词汇研究。他的研究思想是功能语法的核心内容,也是弗斯语言学理念的体现。

弗斯的同事辛克莱是语料库语言学的主要代表和推动者,也是新弗斯学派的领军人物。早在 1966 年他就提出了基于语料库进行词汇和语法研究的具体规划。随后,他主持建成了目前世界上最大的语料库——英语银行(Bank of English)。借助该语料库,他在上个世纪 80 年代提出用语法形式和词汇相结合的方法来描述英语。新弗斯学派延续了弗斯的语言学理论,

并开创了功能强大的语料库研究方法,提出了许多新的语言学理念和方法论,深化了语料库研究的内涵,推动了语料库语言学的发展。

1.2.2 源于结构主义的理论依据

美国结构主义语言学 (Structuralist Linguistics) 又称美国描写语言学 (Descriptive Linguistics), 布龙菲尔德是该学派的主要代表人物。作为一种科学研究方法, 结构主义学派主张实证研究, 认为只有不同的人都能看得见的现象才是证明或反驳一种理论的理想材料。这一观点为以后语言学描述从单纯的定性化发展成为更真实可靠的定量评价打下了基础。同弗斯语言学一样, 结构主义也反对语言存在于集体心智当中的观点。

就语言研究自身而言, 布龙菲尔德 (1933) 注重研究的科学性和客观性, 他把经验方法和直接观察到的事物和现象等同起来, 认为科学的语言研究方法必须以自然的语料为基础。

当前的语料库语言学主要利用语料库进行语言学某个层面的研究, 是一种新的研究手段, 缺乏完善的理论支撑。但是, 通过前文的回顾我们可以看到它的很多方法除了受到经验主义及弗斯学派的影响外, 在很大程度上还受到了结构主义学派的左右。例如, 语料库语言学强调采用的概率法, 以实际使用中的语言现象的出现概率为语言研究的主要依据, 从定量的角度为语言规律的描述提供科学可靠的数据支撑, 这同结构主义所倡导的实证研究导向基本吻合。

语料库语言学的理念、方法和原则大都基于语料的采集与分析, 这与理性主义、弗斯语言学和美国结构主义语言学采用的研究方法和研究范式存在很多相同之处。这也可以说是对此前语言学理论的一种融合、提升和发展。即使在今天, 弗斯和布龙菲尔德的许多观点和主张在指导语料库建设和研究时依然具有重要的理论和实践价值。很多学者继承和发扬了这两个学派的主要思想和观点, 坚持以语言意义的研究为主导, 开拓了一系列的语料库研究方法, 提出了很多全新的学术理念和研究策略。这些指导思想在很大程度上深化了语料库语言学的内涵, 奠定了语料库语言学的理论基础, 同时引领和昭示着其未来的发展方向。

1.3 语料库语言学的本质

长期以来, 语言研究者们对语料库语言学的本质一直争论不休。语料库语言学到底是一种新的研究方法还是新的学科? 它与现代语言学各主要流派的关系如何? 关于这些问题, 主要有以下两种观点:

第一种观点将语料库研究视为一种研究方法, 认为语料库语言学是一种可以应用于任何语言研究领域的方法论 (McEnery & Wilson, 2001), 容易与其他语言分支结合 (G. Leech, 1992), 但是它在目前的研究水平下, 还称不上是一种完备的理论或一门成熟的学科, 仅仅是

一种新的研究手段 (S. Hunston, 2006; G. Kennedy, 2000; E. Tognini-Bonelli, 2001)。作为一种理论架构,语料库语言学使我们对语言的理解产生一种质的改变 (Halliday, 1991),因此它也是一种全新的哲学思维方式,是一种正受到越来越广泛注意和接受的语言观 (J. Svartvik, 1996)。

第二种观点认为,英语语料库语言学已经在语言学界确立了自己的地位,已成为一门独立的、新兴的语言学分支学科”(刘满堂,2004:98),它是“当今最具变革意义的发展”,它能够依据新的语言事实、描述手段和描述系统对结构、意义和功能进行全新的阐述,这是学科属性的突出表现 (卫乃兴,2009:41),同时它还拥有完整的理论体系,独特的研究方法,所以其学科地位是毋庸置疑的 (李文中,2010)。

综上所述,我们认为,借助于语料库的数据检索,语言学家可以为他们的理论假设寻求有力的数据验证,可以推演出一系列句法、语义、语篇和语用规则。由于计算机标注和检索工具的发展,语料库的功能越发强大,基于语料库的研究视野也更加开阔和纵深。在这个意义上,语料库语言学已经成为当代语言研究过程中不可替代的工具和方法。但是如果把它视为完善的学科体系,我们认为还不够成熟。纵观国内外的语料库语言学相关书籍和文献,绝大部分是关于语料库驱动的或者基于语料库的应用性研究。相反,有关语料库语言学的本质、理论基础、研究模式和研究体系的论断和思考却十分罕见。如果语料库语言学成为独立的学科,这需要更多的语言学家和语料库爱好者投注更多的精力、做更深层次的挖掘,提出普遍接受的理论。我们也相信,这一天终会来临。

1.4 语料库语言学的研究特点

1.4.1 语言研究的特点

D. 拜伯 (D. Biber) (2000) 认为,语言研究可分为两大领域:语言结构与语言用法。语言用法的研究着重关注语言在文本中的实际使用,即特定语境下,如何恰当有效地使用语言,实现成功的沟通。语言结构研究涉及下列内容:探讨相似的语言结构在不同语境或文本中的使用效果和实现功能,比如如何表达个人的意图,如何理解说话人的话语目的等;关注说话人或文本作者因文化背景、教育环境和社会经济不同而形成的不同话语风格,以及构成相似意义和语法功能的原因;基于不同文本中的语言变化,研究同一种语言的不同变体的变化特性;揭示语言使用的典型模式。

1.4.2 语料库语言学的研究特点

语料库语言学是以语料库为手段来研究语言,是通过计算机处理的大量语言信息的集合。目前语料库的规模日益扩大,可以储存和处理亿万字符的语料。另外借助赋码手段还可以得

出描述语言特征的更详细信息,并依据索引行和联结模式对相关的语言理论进行验证。

拜伯(2000)还指出,语料库语言学主要通过检索相关的“联结模式”(association patterns)来研究语言使用的特征。他还强调,分析语境因素不能依靠直觉或有限的例子,必须分析语言使用者提供的大量和真实的语料。他认为语料库语言学能处理和加工大量语料,从中提取有价值的语言学数据,是语言用法研究的有力工具,它的研究方法和模式为语言学研究开辟了新的道路。语料库语言学能够把以分析语言特征和联结程度的定量分析同强调功能描述的定性分析有机地结合起来,形成语言使用的实际模式。语料库语言学的实验性与计算机的自动与互动技术紧密相连,没有先进的语料库检索软件和系统很难发挥语料库语言学的强大功能。

亨斯顿(Hunston)(2006)指出,词语索引行是语料库信息处理的最重要的途径,语料库的主要功能在于借助自然取样的数据帮助语言研究者克服和弥补内省法给数据分析带来的不足,从而获得基于大量语言证据而产生的实证结果。

除此之外,李文中在《语料库语言学的研究视野》一文中集中地探讨了语料库语言学的目的、方法和立场。他认为语料库语言学的目的是“研究语言意义”,在语料库语言学研究,“文本是唯一的数据源”。他还认为,语料库语言学的主要方法是“基于语料库(Corpus-based)或语料库驱动(Corpus-driven)的研究”(李文中,2010:37-39)。

卫乃兴则认为,“采取自下而上的研究方法是语料库语言学区别于其他语言学分支最主要的方法特征”,“语料库语言学不排斥解释,但一切解释都需建立在对事实的科学描写基础上”(卫乃兴,2009:36)。

综上所述,国内外学者对语料库语言学的研究特点虽然态度各异,但究其内涵都体现了如下思想:语料库语言学是以意义研究为出发点的描写性语言研究,其研究基础来自于自然真实的语言;通过语料库汇集大规模的语言样本,利用计算机科学技术以及数理统计方法实现对真实语言的定量与定性相结合的科学的客观的分析,力图展现复杂的语言现象背后的简单规律。

1.5 语料库语言学的研究领域

1.5.1 语言研究

语料库是对语言进行定量分析的工具和方法。研究人员按照一定目的和标准搜集或选择一系列语言材料,这些语言材料需具有极强的代表性和规模性。基于这样的标准形成的语料库可以包含海量的语言样本,具有广泛的语体特征;基于这样的语料库可以使语言研究从传统的直觉经验方法转向量化的统计方法,提高了语言研究的效率和准确性。例如,可以借助计算机的强大的运算和信息处理能力对语料库进行自动检索,形成以所要研究的词或短语为中心的索引行,为语言学家进一步的文本分析提供证据。