

语料库翻译学文库

国家社科基金项目成果

上海市十二五规划重点图书

英汉翻译中的 汉语译文语料库研究

肖忠华 著



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

语料库翻译学文库

国家社科基金项目(项目批准号 07BYY011)最终研究成果

上海市十二五规划重点图书

英汉翻译中的 汉语译文语料库研究

**Corpus-Based Studies of Translational Chinese
in English-Chinese Translation**

上海交通大学出版社

内 容 提 要

书中所展示的是国家社科基金项目“基于语料库的英汉翻译语言特征量化研究”(项目批准号 07BYY011)的最终研究成果。该项目将语言对比与翻译研究有机地结合起来,采用基于语料库的研究方法,对汉语母语语料库以及与之对应的汉语译文语料库进行定性与定量的对比分析,从多种角度、在不同层面对汉语译文的特征进行全方位的宏观与微观考察,找出汉语译文的区别性特征,并通过考察英汉平行语料库中的源语渗透效应来确定翻译过程中源语对译文的影响程度。再将针对汉语译文的研究发现与目前国际上现有的主要是基于英文译文的翻译共性研究结果进行对比分析,探讨英汉语这样大跨度的翻译语对中产生的证据对所谓“第三语码”的启示。本书可作为对语料库语言学、翻译学 and 对比语言学感兴趣的学者与研究生的参考用书。

图书在版编目(CIP)数据

英汉翻译中的汉语译文语料库研究/肖忠华著.

—上海:上海交通大学出版社,2012

(语料库翻译学文库)

ISBN 978-7-313-08443-9

I. 英... II. 肖... III. 英语—翻译—研究

IV. H315.9

中国版本图书馆 CIP 数据核字(2012)第 087463 号

英汉翻译中的汉语译文语料库研究

肖忠华 著

上海交通大学出版社出版发行

(上海市番禺路 951 号 邮政编码 200030)

电话:64071208 出版人:韩建民

常熟市文化印刷有限公司印刷 全国新华书店经销

开本:787mm×960mm 1/16 印张:13.25 字数:213 千字

2012 年 6 月第 1 版 2012 年 6 月第 1 次印刷

印数:1~2 030

ISBN 978-7-313-08443-9/H 定价:40.00 元

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系

联系电话:021-65982320

语料库翻译学文库

编 委 会

主编 王克非 胡开宝

编委(以姓氏拼音为序)

桂诗春	胡开宝	胡显耀	黄立波
李德超	李德凤	李德明	李文中
廖七一	刘泽权	秦洪武	王克非
肖忠华	杨承淑	朱纯深	

语料库翻译学文库

总序

语料库翻译学,即基于语料库的翻译研究;它在研究方法上以语言学理论为指导,以概率和统计为手段,以双语真实语料为对象,对翻译进行历时或共时的研究,代表一种新的研究范式。语料库翻译学一般依靠双语语料库,主要是翻译语料库(translational corpus)、平行/对应语料库(parallel corpus)和类比语料库(comparable corpora),并需分别进行不同的标注。因此对于语料库翻译学的发展来说,有两大关键点:一是双语库的研制,它是技术手段,是基础设施,也体现研制者的研究目的;其加工方式和程度直接影响研究课题的开展。二是研究切入点,即利用双语库能给翻译研究带来什么新发现、新解释,它关系到语料库翻译学如何解决翻译上的各种问题,或如何深化人们对翻译现象和活动的认识。

本丛书为国内(迄今为止也是国际上)第一套基于语料库的翻译研究系列专著,主要选题有“语料库翻译学探索”、“语料库翻译学概论”、“类比语料库及其研究”、“基于语料库的翻译语言特征研究”、“基于语料库的翻译文体研究”、“基于语料库的翻译搭配研究”、“莎士比亚戏剧英汉平行语料库的创建与应用研究”、“语料库与翻译教学”等。其最大特点有三。一是研究者均为国内本领域专家,著作为国家和省部级科研项目成果;二是所收专著各有专攻,覆盖平行语料库、翻译语料库和类比语料库三方面研究;三是成果为第一批与中文相关的语料库翻译学文献,并伴有一批语料库的建立和充实。

语料库翻译学逐步从方法论发展成为连贯、综合、丰富的范式,应用于翻译理论以及翻译的描写和实践等一系列问题的探讨。新的研究范式带来了研究思路的更新和研究重点的转移。编者和作者期盼这套丛书能促进我国的翻译和语料库两方面学术研究的更大发展。

王克非

2011年夏书于北外家中

序

我第一次见到肖忠华博士是五六年前的的事了。当时我在香港中文大学翻译系任教,他应邀到英文系访问并讲座。由于先前读过他的论文,并有朋友推荐说“忠华人很好”,所以收到英文系发来的有关讲座的海报,便带了我的博士生一同前往。讲座的题目是关于语料库语言对比研究的。虽然准确题目已经记不太清楚了,但对忠华博士对语料库研究的热忱和学识印象深刻。其后,我离开香港,赴英国伦敦大学任职。由于地理距离近了,更重要的是由于对语料库语言对比与翻译研究的共同兴趣,与忠华多有接触并一直保持联系。近日收到他通过网络发来的专著《英汉翻译中的汉语译文语料库研究》书稿,并嘱咐为其作序。

其实,忠华博士真的无需推介。他师从语料库语言学大家、英国兰卡斯特大学 McEnergy 教授,从博士研究开始一直潜心研究语料库语言对比及翻译,在英国主持和承担了多项与语料库相关的科研项目,英文著述甚丰。如今,忠华博士也已经是国际语料库语言对比研究方面的专家了。所以,下面几段文字只是我阅读这部书稿时的一点随想和体会。

自 20 世纪下半叶,翻译学作为一门新兴学科取得了迅猛的发展。个人认为,其间翻译学发生了两次重大革命。其一是研究导向的变革,翻译学从规范性研究传统转向了描写翻译研究,从以前的翻译研究必须为翻译实践服务,扩大到更多地关注翻译的动因、目的、过程、影响翻译的社会文化因素以及翻译作品在译入语文化中的作用和影响等。至此,研究的视野开阔了,翻译学加强了与其他相邻学科的交叉融合,并由此催生了一系列的描写翻译学理论,如翻译的多元系统论与操控论、女性主义翻译理论、后殖民翻译理论以及近来兴起的社会学的翻译观等。另一次革命是翻译学研究方法论的革新。翻译研究者渐渐抛弃了个人翻译经验总结式的翻译研究,转而追求翻译研究方法的客观与严谨。随之产生的就是基于实证研究方法的口译研究、翻译的心理认知过程研究、翻译教学研究等等。而语料库翻译研究可謂是翻译学两次大革命的共同成果,即以科学实证的研究方法,聚焦翻译作品,对翻

译语言的特征进行描写,并试图洞察翻译的过程,尽可能地对翻译过程中所发生的事情进行解读;或者描述对比译者的风格,并力图寻找其风格背后形成的原因。

语料库翻译研究发轫于 20 世纪 90 年代。一般以英国曼彻斯特大学翻译学教授贝克(Mona Baker)博士 1993 年发表的论文《语料库语言学与翻译研究:启示和应用》为标志。该文一经发表立即引起了翻译研究界的关注。以贝氏指导的博士生为中坚力量的一批学者采用语料库的方法,从不同的角度、针对不同的翻译问题展开研究。概括而言,这些研究主要沿着四个维度展开,即论证关于“翻译共性”的假设、对比译者风格、描述翻译语言特征以及探究语料库在翻译教学中的运用。通过将近 20 年的努力,发表和出版了一批非常有趣、价值极高的论文和专著。

中国近年来在该领域的发展速度及其取得的成绩也十分令人振奋。自 2009 年起,无论是相关论文的数量,还是相关的学术会议数量,特别是从事相关研究的学者和学生人数,都远远超过了任何其他国家。当然,美中不足的就是,相当一部分的研究仍然停留在模仿前人的研究,较少开拓;研究设计和方法也多仿照他人,缺乏创新,自然研究结果和结论也就难有突破。鉴于此,忠华博士的这部以中文撰写的语料库研究专著就具有特别重要的意义。

作者采用语料库的方法,首先将汉语译文语料库与对应的汉语母语语料库进行对比分析,从不同的角度和层面对汉语翻译文本的特征进行全方位考察,确定汉语译文的区别性特征。其次将汉语译文语料库与英语原文语料库进行平行对比,考察英汉翻译中的源语渗透效应,并以此来确定翻译过程中源语对译文的影响。其研究设计巧妙,论证严谨充分。所以,在一定程度上讲,其关于汉语译文体特征的描述具有结论性的意义。

然而,这项研究最大的特点还是其在研究方法上的突破。正如作者在书中所述,目前国际上语料库研究使用较多的是单一的方法,即通过类比语料库对比译文体和母语文本的语言特征,或者使用双语平行语料库研究原文与译文之间的对应关系。而忠华的研究采用的是复合研究方法,即结合使用类比语料库和平行语料库,从不同角度对译文进行考察,系统地探索汉语译文的多维度特征。所以,这本专著也完全可以当作一部研究方法论的教材,指导以语料库方法研究语言对比。而这样的方法论教材正是中国语料库研究界所缺少的。我相信,任何有志从事语料库语言对比或翻译研究的研究生和青年教师,如果能认真仔细研读本书,必定会获益匪浅,并在今后的研究中有所创新和突破。

英国伦敦大学 亚非学院

李德凤

2012 年 5 月 16 日

鸣 谢

在本书付梓出版之际,笔者想利用这个机会首先向国家社科基金多年来对本研究课题(编号 07BYY011)的资助表示感谢。其次要感谢的是“语料库翻译学文库”的两位主编,王克非教授和胡开宝教授,以及上海交通大学出版社管新潮副编审,感谢他们对文库的策划和对本书出版的大力支持。同时,笔者还要对项目组成员的辛勤付出表示感谢。浙江大学的何莲珍教授和乐明副教授为浙大汉语译文语料库 ZCTC 的研制以及本项目发起的首届“基于语料库的语言对比与翻译研究”(UCCTS)国际学术研讨会做出了很大的贡献,并感谢她们多年来在项目管理方面提供的支持。福建工程学院/澳门大学的戴光荣老师参与了本项目的后期研究,他在语料分析和阶段性成果的撰写方面提供了大力协助。浙江大学外语学院的 2008 级研究生黄伟、于书林、胡文颖、闵尚超等为语料收集工作付出了辛勤的劳动。最后要感谢的是嘉兴学院的贺文照副教授,他为本书第 3 章中有关汉语译文研究现状的综述收集了许多资料。没有他们的辛勤付出和大力支持,笔者将无法顺利完成本书的撰写。

肖忠华

2012 年 3 月于杭州

目 录

第 1 章	绪论	1
1.1	翻译研究中的范式转变	2
1.2	研究目的及意义	4
1.3	本书框架结构	5
第 2 章	语料库翻译学:翻译研究的新范式	7
2.1	语言研究中的语料库“革命”	7
2.2	用于语言对比与翻译研究的语料库类型	9
2.3	基于语料库的翻译研究	12
第 3 章	翻译语言研究现状	20
3.1	有关翻译共性假设的研究	20
3.2	汉语译文研究现状概述	23
3.3	汉语译文语言特征研究	25
3.4	现有研究中存在的问题	30
第 4 章	语料库及检索工具	33
4.1	现有主要通用汉语语料库简介	33
4.2	兰卡斯特汉语语料库(LCMC)	35
4.3	浙江大学汉语译文语料库	42
4.4	本研究使用的平行语料库	47

4.5	语料分析与统计工具	48
第5章	汉语译文的宏观统计特征	62
5.1	词汇密度与文本信息密度	62
5.2	词频分析	66
5.3	平均句子长度与平均段落长度	70
5.4	LCMC与ZCTC中的词簇	73
第6章	汉语译文的词汇特征	81
6.1	LCMC与ZCTC中的关键词与关键词类	81
6.2	LCMC与ZCTC中的词类分布与词类簇	84
6.3	LCMC与ZCTC中标点符号的分布	92
6.4	LCMC与ZCTC中的代词	94
6.5	LCMC与ZCTC中的连接词	97
6.6	LCMC与ZCTC中的习语	100
6.7	LCMC与ZCTC中的话语重述标记	103
第7章	汉语译文的语法特征	114
7.1	LCMC与ZCTC中的“被”字句	114
7.2	LCMC与ZCTC中的“把”字句	124
7.3	LCMC与ZCTC中的“有”字句	127
7.4	LCMC与ZCTC中的“是”字句	129
7.5	LCMC与ZCTC中的“所”字句	133
7.6	LCMC与ZCTC中的“连”字句	135
7.7	LCMC与ZCTC中的量词	136
7.8	LCMC与ZCTC中的体标记	139
7.9	LCMC与ZCTC中的结构助词	142
7.10	LCMC与ZCTC中的语气词	145

第 8 章 汉语译文特征对翻译共性研究的启示	147
8.1 汉语译文的宏观统计特征	147
8.2 汉语译文的词汇特征	150
8.3 汉语译文的语法特征	153
第 9 章 结论	157
9.1 翻译体汉语视角下的翻译共性假设	157
9.2 本研究的突破及其对语料库翻译研究的贡献	160
9.3 本领域将来的研究方向	162
附录 1 LCMC 与 ZCTC 语料库词类标注集	164
附录 2 浙大汉语译文语料库 ZCTC 中的关键词	167
附录 3 本书随附语料库说明	169
参考文献	170

图 示 目 录

图 4.1	LCMC 语料库文件头	40
图 4.2	LCMC 段落标注示例	41
图 4.3	ZCTC 语料库文件头	44
图 4.4	ZCTC 段落标注示例	45
图 4.5	Xaira 索引工具	50
图 4.6	Xaira 用户终端程序工具栏(1)	52
图 4.7	Xaira 用户终端程序工具栏(2)	52
图 4.8	WordSmith 检索功能检索窗	54
图 4.9	WordSmith 检索功能词簇窗	54
图 4.10	WordSmith 检索功能词簇参数设置	55
图 4.11	WordSmith 索引基本参数设置	55
图 4.12	WordSmith 索引关系参数设置	56
图 4.13	WordSmith 词表功能	56
图 4.14	WordSmith 索引结果	57
图 4.15	WordSmith 词簇设置选项	57
图 4.16	英国英语语料库 FLOB 中的 3 词词簇	58
图 4.17	WordSmith 关键词设置	59
图 4.18	ParaConc 平行检索示例	60
图 5.1	LCMC 与 ZCTC 中的词汇密度	63
图 5.2	LCMC 与 ZCTC 中的标准型次比	64
图 5.3	非文学类语体(A-J)和文学类语体(K-R)中的标准型次比	65

图 5.4	LCMC 和 ZCTC 中实义词与功能词之比	65
图 5.5	LCMC 与 ZCTC 中的高频词	68
图 5.6	LCMC 与 ZCTC 中的罕用词	68
图 5.7	LCMC 与 ZCTC 中的平均词长	69
图 5.8	LCMC 与 ZCTC 中的平均句长	70
图 5.9	LCMC 与 ZCTC 中的平均句子片段长度	72
图 5.10	LCMC 与 ZCTC 中的平均段落长度	73
图 5.11	LCMC、ZCTC 和 FLOB 中的 2~4 词词簇	75
图 5.12	LCMC、ZCTC 和 FLOB 中的高频词簇	76
图 5.13	LCMC、ZCTC 与 FLOB 中 2 词词簇的覆盖率分布	77
图 5.14	LCMC、ZCTC 与 FLOB 中 3 词词簇的覆盖率分布	77
图 5.15	LCMC、ZCTC 与 FLOB 中高频高覆盖率词簇的语体分布	78
图 6.1	LCMC 与 ZCTC 中的实义词分布	84
图 6.2	LCMC 与 ZCTC 中的功能词分布	85
图 6.3	LCMC 与 ZCTC 中标点符号的总体分布	92
图 6.4	LCMC 与 ZCTC 中主要标点符号的分布	93
图 6.5	LCMC 与 ZCTC 中代词的总体分布	94
图 6.6	LCMC 与 ZCTC 中的主要代词种类	95
图 6.7	LCMC 与 ZCTC 中连接词的归一化频率	97
图 6.8	LCMC 与 ZCTC 中不同频段连词的分布	98
图 6.9	正式用法连接词在 LCMC 与 ZCTC 中的分布	99
图 6.10	非正式用法连接词在 LCMC 与 ZCTC 中的分布	99
图 6.11	LCMC 与 ZCTC 中习语的归一化频率	102
图 6.12	LCMC 与 ZCTC 中书面语化与口语化的话语重述标记	107
图 6.13	书面语化的话语重述标记在汉语各语体中的分布	108
图 6.14	口语化的话语重述标记在汉语各语体中的分布	109
图 7.1	短式与长式被动句的分布	115
图 7.2	被动句式的语用含义分布	116
图 7.3	LCMC 与 ZCTC 中“被”字句的分布	117

图 7.4	Babel 英汉平行语料库中“被”字句检索截图	119
图 7.5	英译汉文学与非文学语料中被动句式的源语渗透效应	122
图 7.6	LCMC 与 ZCTC 中“把”字句的归一化频率	126
图 7.7	LCMC 与 ZCTC 中“有”字句的归一化频率	129
图 7.8	LCMC 与 ZCTC 中“是”字句的分布	130
图 7.9	LCMC 与 ZCTC 中“所”字句的分布	133
图 7.10	LCMC 与 ZCTC 中“连”字句的分布	135
图 7.11	LCMC 与 ZCTC 中量词的归一化频率	137
图 7.12	LCMC 与 ZCTC 中量词分布的语体差异	138
图 7.13	LCMC 与 ZCTC 中完整体标记“了”的归一化频率	140
图 7.14	LCMC 与 ZCTC 中完整体标记“过”的归一化频率	141
图 7.15	LCMC 与 ZCTC 中非完整体标记“着”的归一化频率	142
图 7.16	LCMC 与 ZCTC 中助词“的、地、得”的归一化频率	143
图 7.17	LCMC 与 ZCTC 中助词“之”的归一化频率	144
图 7.18	LCMC 与 ZCTC 中语气词的归一化频率	145

表 格 目 录

表 4.1	LCMC 语料库选材范围	37
表 4.2	LCMC 语料库抽样年代	38
表 4.3	LCMC 语料库文本的 XML 标注	40
表 4.4	LCMC 与 ZCTC 语料库设计	43
表 4.5	ICTCLAS 词性标注第一层次的词性类别	46
表 5.1	不同语体词汇密度的平均差及统计分析	63
表 5.2	各文本类别中实义词与功能词之比	66
表 5.3	LCMC 与 ZCTC 中的词频统计	67
表 5.4	LCMC 与 ZCTC 中平均词长百分比	69
表 5.5	LCMC 与 ZCTC 中的平均句子片段长度	72
表 5.6	LCMC 与 ZCTC 中不同词簇的频率及其对数似然测试	75
表 5.7	汉语译文与母语中的高频词簇	79
表 5.8	部分 3~6 词词簇示例	80
表 6.1	ZCTC 中的关键词	82
表 6.2	汉语译文语料库 ZCTC 中的关键词类	83
表 6.3	LCMC 与 ZCTC 中实义词对比	85
表 6.4	LCMC 与 ZCTC 中的功能词对比	86
表 6.5	LCMC 中最常用 3 词词类簇及其在 ZCTC 中的排序	87
表 6.6	ZCTC 中最常用 3 词词类簇及其在 LCMC 中的排序	88
表 6.7	在 LCMC 中出现而 ZCTC 中没有出现的高频高覆盖率 3 词词类簇	89

表 6.8	在 ZCTC 中出现而 LCMC 中没有出现的高频高覆盖率 3 词词类簇	90
表 6.9	ZCTC 中的关键词类簇与负关键词类簇	91
表 6.10	LCMC 与 ZCTC 中主要标点符号的对数似然测试	93
表 6.11	LCMC 与 ZCTC 中代词分布的对数似然测试	95
表 6.12	LCMC 与 ZCTC 中主要代词分布的对数似然测试	96
表 6.13	LCMC 与 ZCTC 中习语使用频率及其对数似然测试	102
表 6.14	Chen(2006:152)中常用话语重述标记在不同语料库中的分布	105
表 6.15	LCMC 与 ZCTC 中书面语化和口语化的话语重述标记	106
表 6.16	LCMC 与 ZCTC 中的话语重述标记分布	108
表 6.17	英语中的常用话语重述标记	110
表 6.18	英语中口语化及书面语化的话语重述标记分布	110
表 6.19	英汉翻译中的话语重述标记明晰化	111
表 7.1	LCMC 与 ZCTC 中被动句式的对数似然测试	117
表 7.2	LCMC 与 ZCTC 中“把”字句的对数似然测试	126
表 7.3	LCMC 与 ZCTC 中“有”字句的对数似然测试	129
表 7.4	LCMC 与 ZCTC 中“是”字句的对数似然测试	131
表 7.5	LCMC 与 ZCTC 中“所”字句的对数似然测试	134
表 7.6	LCMC 与 ZCTC 中“连”字句的对数似然测试	136
表 7.7	LCMC 与 ZCTC 中量词的对数似然测试	137
表 7.8	LCMC 与 ZCTC 中量词分布的对数似然测试	138
表 7.9	LCMC 与 ZCTC 中“了”的对数似然测试	141
表 7.10	LCMC 与 ZCTC 中“过”的对数似然测试	141
表 7.11	LCMC 与 ZCTC 中“着”的对数似然测试	142
表 7.12	LCMC 与 ZCTC 中助词“的、地、得”的对数似然测试	144
表 7.13	LCMC 与 ZCTC 中助词“之”的对数似然测试	144
表 7.14	LCMC 与 ZCTC 中语气词的对数似然测试	146

第1章

绪 论

本书研究对象为英汉翻译中汉语译文相对于汉语母语而言所呈现出来的语言特征,属描写翻译研究中翻译共性的研究范畴。“翻译共性”(translation universal)这一概念由 Baker(1993:243)首先提出。她将其定义为“翻译文本而不是源语文本中出现的典型语言特征,这些特征不是特定语言系统干扰的结果。”也就是说,翻译共性是指与源语或译入语母语相比,翻译体语言所表现出来的规律性语言特征。源语对于翻译的影响效果非常明显,翻译语言特征也不同于译入语母语特征(McEnery & Xiao 2007a)。这种翻译语言的区别性特征可以通过比较翻译语料和译入语母语中的对应语料来发现。此类研究既可以为翻译过程提供参考借鉴,同时也有助于对翻译规范或所谓的“第三语码”(the third code, Frawley 1984)的理解。

书中所展示的是国家社科基金项目“基于语料库的英汉翻译语言特征量化研究”(项目批准号 07BYY011)的最终研究成果。该项目将语言对比与翻译研究有机地结合起来,采用基于语料库的研究方法,对汉语母语语料库以及与之对应的汉语译文语料库进行定性与定量的对比分析,从多种角度、在不同层面对汉语译文的特征进行全方位的宏观与微观考察,找出汉语译文的区别性特征,并通过考察英汉平行语料库中的源语渗透效应来确定翻译过程中源语对译文的影响程度。再将针对汉语译文的研究发现与目前国际上现有的主要是基于英文译文的翻译共性研究结果进行对比分析,探讨英汉语这样大跨度的翻译语对中产生的证据对所谓“第三语码”的启示。

本章首先简要地回顾过去半个多世纪以来翻译研究中的范式转变,以提供必要的研究背景,然后讨论本课题的研究目的及意义,最后介绍本书的框架结构。