

语料库翻译学文库

# 基于新型 历时复合语料库的 翻译汉语特征研究

庞双子◎著



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

语料库翻译学文库

本书受国家社科青年基金项目资助（项目编号15CYY005），  
为国家社科基金优秀结项成果

# 基于新型 历时复合语料库的 翻译汉语特征研究

庞双子◎著



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

## 内容提要

本书为“语料库翻译学系列”丛书之一,是国家社科青年基金项目成果,结项获评优秀。本书通过构建新型语料库,运用多元统计的方法分别考察了20世纪30年代、60年代、90年代三个时间段的文学翻译文本在语体、词汇和语法上的语言表征,并对英语源语文本、汉语翻译文本、汉语原创文本以及未受翻译影响的汉语原生态文本之间的相互关系和变化进行了系统的、连续的考察,进一步探究了翻译与原创汉语语言发展间的关系。本书为翻译与目标语语言发展的考察提供了新的研究路径,适合各高等院校的语言学教师以及硕士和博士研究生等学习使用。

## 图书在版编目(CIP)数据

基于新型历时复合语料库的翻译汉语特征研究 / 庞  
双子著. —上海: 上海交通大学出版社, 2021  
ISBN 978-7-313-24023-1

I. ①基… II. ①庞… III. ①英语—翻译—研究  
IV. ①H315.9

中国版本图书馆 CIP 数据核字(2021)第 080423 号

## 基于新型历时复合语料库的翻译汉语特征研究

JIYU XINXING LISHI FUHE YULIAOKU DE FANYI HANYU TEZHENG YANJIU

著 者: 庞双子

出版发行: 上海交通大学出版社

邮政编码: 200030

印 制: 江苏凤凰数码印务有限公司

开 本: 710 mm×1000 mm 1/16

字 数: 213 千字

版 次: 2021 年 6 月第 1 版

书 号: ISBN 978-7-313-24023-1

定 价: 82.00 元

地 址: 上海市番禺路 951 号

电 话: 021-64071208

经 销: 全国新华书店

印 张: 13.75

印 次: 2021 年 6 月第 1 次印刷

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: 025-83657309

# 序

翻译是人类跨语言的交际活动,它使一地一族的文化和思想传到了更广阔的地域空间,也使另一地另一族的文化和思想因交流而丰富,从而收到绽放异彩的效果。所以翻译即新水注入,如新桥架构,似新路铺设,使心心相通,是功德无量的事业。

我上述这段话,含有两层意思。第一层意思是说翻译沟通了不同的语言文化,完成其初步使命;第二层意思是说翻译带入的域外语言、思想、文化可能会对目标语及其文化和社会产生深远的影响,实现其终极使命。翻译语言,即译文的载体,本身是沟通的产物,也是受到影响的产物,因它打上了外语的印记,已不再是本来的目标语。翻译语言是客观存在,它所带来的一些异样的新鲜表述会持久地影响目标语的使用。一百多年来翻译的汉语就是翻译语言的典型代表,但它如何以及多大程度上影响了汉语白话文的成长,长期以来学界未给予足够的关注。

现在是发掘这一研究资源,探究透过翻译的语言接触问题的时候了。语料库给了这类研究极好的助力。

例如摆在我们面前的这本《基于新型历时复合语料库的翻译汉语特征研究》。作者庞双子,本科、硕士毕业于天津外国语学院,博士毕业于北京外国语学院,随后在上海交通大学完成博士后研究。这部著作伴随了她的成长历程。2015年夏天,正在北京外国语学院攻读博士学位的作者获得了国家社科基金青年项目(15CYY005)资助,课题即本书书名。2019年,作者在上海交通大学外国语学院副教授的岗位上,一边忙于教学和研究工作,一边照顾重病的家人,坚韧地完成了这个项目。2020年春末,该项目顺利通过结项评审,且获评优秀(仅6%的优秀率)!五年前获得项目的惊喜,化为五年后完成课题的欣

喜,我想,这是她青年时期最为欣慰和骄傲的事情。

这部书有三大亮点。其一,本书构建了一种新型的语料库,即多时(历时和共时)、多种(平行、类比、参照三库复合)、多域(多个文学文本类型)的语料库,总库容达一千万字词,这是语料库研究上的突破,为翻译汉语的历时考察,翻译对于原创汉语语言的影响提供了有力的数据支撑,为此类研究提供了新的路径。

其二,书中应用上述历时复合语料库,分别选取 20 世纪一百年间三个时间段即 30 年代、60 年代、90 年代的汉语文学翻译文本,考察其在词汇和语法上的表现,寻找翻译汉语的语言特征。例如在语体上,作者通过多维统计方法,将其分为五个维度进行考察。在词汇特征上,考察翻译文本的关键词、高频词、低频词、连接词、代词、话语重述标记、短语、搭配和类联接等。在句式特征上,分析翻译文本中“被”字句、“把”字句、存现句、“是”字句、体标记等的使用。这些都是富有新意的研究。

其三,不止于发现翻译汉语的语言特征,作者还将英语源语文本、汉语翻译文本、汉语原创文本以及未受翻译影响的汉语原生态文本之间的相互关系和变化作系统的、连续的考察,进一步探究翻译与目标语语言发展间的关系,从而发现翻译文本中的部分语言特征会迁移至原创汉语文本,对后者在语体、词汇、句式、语义以及交际功能上产生影响。这是研究的深化。

此外,作者运用当前国际上较为前沿的多元统计方法,包括方差分析、对应分析、随机森林、回归分析等,结合 R 语言自身的统计优势,对翻译文本和原创文本的历时变化进行了量化的考察,大量的数据、图表使这项研究显得更为信实可靠。

放眼国际上的相关研究,会发现绝大多数研究是基于英语与其他欧洲语言之间的翻译文本开展的。我们希望中国的学者补上基于汉语与英语等语言之间的翻译语言特征研究,将中国的语料库翻译学研究推向国际前沿。

这本新著,正是我国这类研究的起步。谨以此短序,祝福我们的青年学者!

王克非 序于北京外国语大学寓所

2021 年 3 月 31 日

# 前 言

20 世纪 90 年代,“翻译共性”的提出使得翻译文本语言特征的实证研究逐渐聚集,并发展成为一片崭新的研究区域。透过翻译语言这条路径可以看到翻译文本不同于原创文本的特征,考察翻译与语言发展之间的关联,进而探究翻译行为的本质。随着有关翻译普遍性研究在国际领域的展开,历时研究的发展和利用语料库进行翻译与语言发展关系的考察在国际上成为一个新兴课题。

提及翻译在语言演变历程中的作用,无论是在欧洲还是我国都有较为深邃的历史积淀。早在公元前 1 世纪,罗马翻译者将希腊语译成拉丁文,模仿希腊语的表达方式丰富本族语言。公元 8~11 世纪,拉丁语的宗教翻译逐渐对英语、德语等语言发生影响。到了 20 世纪 50 年代,随着语言接触研究在现代语言学领域的兴起,翻译对于目标语语言的影响的理论建构逐渐形成。20 世纪 90 年代以来,语料库翻译研究的兴起使得该领域的实证研究成为可能。

在英汉语言研究领域,较早可以追溯到王力先生对于现代汉语欧化现象的探讨。该领域研究大致可以分为两种路径,即较为传统的关于“欧化”的探讨和基于语料库讨论翻译与语言发展之间的关系。经过多年发展和探讨,该领域研究的瓶颈有二:一则判断翻译文本中的语言特征如何渗透至原创文本;二则翻译和语言接触这两个因素扑朔迷离,很难界定一些语言特征究竟是由翻译因素所致,还是由于语言接触因素所致。在研究方法上,有关历时考察的研究往往是将某个特定时期的共时语料库进行拼接,其研究结果缺乏可复制性,且同类研究之间验证困难。大型且内部统一的历时语料库亟待建设,相关的统计方法也相对滞后。

上述正是本书所要探究的问题。通过构建新型历时复合语料库,即平行、

类比、参照三者复合的历时语料库,分别考察 20 世纪 30 年代、60 年代、90 年代三个时间段的文学翻译文本在词汇、语法以及部分语义层面上的语言表征,将英语源语文本、汉语翻译文本、汉语原创文本以及未受翻译影响的汉语原生态文本之间的相互关系和变化作系统且连续的考察,深入探究翻译与目标语语言发展间的关系。

本书基于以上历时复合型的语料库,纵向观察了翻译语言词汇特征和语法特征的变化,具体包括语体、词汇、句式和显化特征四个层面。通过质性研究和量化考察相结合,使用多方验证的方法对假设进行效度检验。并将语料库驱动的方法应用其中,结合平行、类比及参照语料库,较为系统地观察源语、译文及原创汉语的发展变化,分析三者之间的关系和相互影响。

本书的贡献主要体现在四个方面:一是语料库方面。本书研制和加工了翻译体汉语和原创汉语一百年间三个时段的新型历时复合型的语料库。这种语料库的建设为翻译汉语的历时考察,翻译对于原创汉语语言的影响提供了有力的数据支撑和实证考察,为该领域的研究打开了新的研究路径。

二是理论方面,本书将历史语言学与翻译研究结合。借鉴目前国际上新兴的透过翻译的语言接触理论,既深入考察英语(源语)对翻译汉语的影响,也关注译语对原创汉语(原生汉语)的间接影响,分析二者的关联及影响机制。将历史语言学与翻译研究相结合,是对语言发展史研究的一个贡献。

三是拓宽第三语码研究。对翻译中显化现象的历史性考察使我们不再停留在目前有限的动因阐释上面,这种加入意识形态因素的考察可以使我们更深刻地认识到显化特征的本质,也对长期以来在语料库翻译学领域争论不休的共性问题做出了历时维度上的诠释。

四是使用多维统计方式。本书通过基于 R 语言可视化的多元统计,对翻译语言特征进行了类似深度 X 光片的捕捉,由此将此项研究推向较深层次。本书对于翻译文本和原创汉语文本两类文本的历时变化进行了量化的考察,并通过方差分析、指标预测、多元回归、随机森林等多种方式对文本进行数据挖掘。将翻译文本中的指标和原创汉语文本中各维度指标结合起来进行效应分析,是以往尚未开展过的研究。

本书的框架结构为:第 1 章绪论,介绍研究背景、思路、方法和价值,主要通过历时多样语料库的构建探讨翻译文本特征的特点以及翻译与语言发展的

关系。接下来第2章通过文献计量的方法对翻译与语言接触研究进行梳理,重点介绍该领域基于语料库的实证研究进展。第3章介绍翻译语言特征和翻译汉语特征研究进展。第4章对语料库的发展现状进行勾勒,并对本书所构建和使用的研究方法,包括新型历时复合语料库的构建、研究步骤、统计方法等进行说明。第5章主要通过平行语料库的方法考察英语源语文本和汉语翻译文本在这三个时期是否具备相关性,同时考察源语透过性的影响。主要包括语体考察、词汇特征、句式特征、显化考察四个方面。第6章通过类比语料考察翻译文本和原创文本两类文本在以上指标的相关性和历时变化,并通过比较部分指标和参照库来进行补正。第7章在前两章实证分析的基础上总结源语透过性变化和译语透过性变化,并讨论该结论对于翻译共性、语言接触理论以及语言发展变化的启示。第8章对本书研究作一总结,并探讨该领域的未来研究走向。

本书从设计、建库、数据分析、数据阐述到研究发现,历时近五个年头,书中的多数章节是笔者在读博期间以及在博后研究工作期间完成。在本书出版之际,特别感谢笔者的博士导师王克非教授对该研究诸环节给予的倾心指导、意见和建议,使得该书可以顺利成稿。感谢上海外国语大学语料库研究院院长胡开宝教授对笔者在站期间的指导、帮助,以及鼓励。感谢上海交通大学外国语学院领导与同事的关心和支持。还要非常感谢国家社科基金的资助,感谢国家社科办匿名评审专家给予的高度肯定和修改建议,这也将是笔者继续在学术之路上进步的莫大动力。

2021年3月8日

庞双子 于上海交通大学

# 目 录

## 第一部分 理论方法篇

### 第 1 章 绪论 3

---

- 1.1 研究背景 / 4
  - 1.1.1 翻译与语言发展关系研究 / 5
  - 1.1.2 翻译语言特征研究 / 6
- 1.2 研究目标 / 7
- 1.3 研究设计 / 8
- 1.4 研究意义 / 9
- 1.5 框架结构 / 10

### 第 2 章 翻译与语言接触研究 11

---

- 2.1 文献计量 / 11
- 2.2 理论溯源 / 16
- 2.3 基于语料库的实证研究进展 / 18

### 第 3 章 翻译语言研究现状 21

---

- 3.1 翻译语言研究进展 / 21

- 3.1.1 从实证研究到理论建构的融合 / 22
- 3.1.2 翻译普遍性本质的拓展和新探 / 22
- 3.1.3 研究方法和统计技术的深入 / 23
- 3.1.4 翻译语言与语言接触研究的契合 / 24
- 3.2 翻译汉语特征的研究脉络 / 26
  - 3.2.1 欧化与杂合 / 26
  - 3.2.2 共时和历时 / 27
  - 3.2.3 实证研究进展 / 28
  - 3.2.4 存在的问题 / 29

## 第4章 研究方法：基于新型语料库

---

31

- 4.1 语料库发展现状 / 31
  - 4.1.1 单语历时语料库的发展 / 32
  - 4.1.2 双语语料库的成长 / 35
- 4.2 新型历时复合语料库 / 39
  - 4.2.1 语料库架构和库容 / 40
  - 4.2.2 语料库建设和统计难点 / 41
- 4.3 研究步骤 / 42
- 4.4 统计方法 / 43

## 第二部分 实证分析篇

## 第5章 平行语料库的分析

---

49

- 5.1 语体特征 / 49
  - 5.1.1 拜伯的语体特征指标 / 49
  - 5.1.2 提取原则和方法 / 51

- 5.1.3 宏观考察 1: 语体维度的单独分析 / 52
- 5.1.4 宏观考察 2: 语体维度的整体考察 / 61
- 5.2 词汇特征 / 63
  - 5.2.1 关键词表、高频词和低频词 / 64
  - 5.2.2 三个时期翻译文本的连接词 / 67
  - 5.2.3 三个时期翻译文本的话语重述标记 / 68
  - 5.2.4 三个时期翻译文本的代词 / 69
  - 5.2.5 三个时期翻译文本的短语单位 / 70
  - 5.2.6 三个时期翻译文本的搭配和类联结 / 73
- 5.3 句式特征 / 76
  - 5.3.1 三个时期翻译文本的“被”字句 / 76
  - 5.3.2 三个时期翻译文本的“把”字句 / 77
  - 5.3.3 三个时期翻译文本的存现句 / 78
  - 5.3.4 三个时期翻译文本的“是”字句 / 78
  - 5.3.5 三个时期翻译文本的体标记 / 79
- 5.4 显化层面 / 79
  - 5.4.1 结构显化 / 80
  - 5.4.2 语义显化 / 100

## 第 6 章 类比语料库和参照语料库的分析

105

- 6.1 语体特征 / 105
  - 6.1.1 宏观考察 1: 语体维度的单独分析 / 105
  - 6.1.2 宏观分析 2: 语体维度的综合考察 / 114
  - 6.1.3 译语透过性中维度与语体指标效应考察 / 115
  - 6.1.4 小结 / 124
- 6.2 词汇特征 / 125
  - 6.2.1 关键词、超用词和少用词 / 125
  - 6.2.2 三个时期翻译文本和原创文本的连接词 / 128

- 6.2.3 三个时期翻译文本和原创文本的话语重述标记 / 129
- 6.2.4 三个时期翻译文本和原创文本的代词 / 130
- 6.2.5 三个时期翻译文本和原创文本的短语单位 / 131
- 6.2.6 三个时期翻译文本和原创文本的搭配和类联结 / 133
- 6.3 句式特征 / 134
  - 6.3.1 三个时期翻译文本和原创文本的“被”字句 / 134
  - 6.3.2 三个时期翻译文本和原创文本的“把”字句 / 135
  - 6.3.3 三个时期翻译文本和原创文本的存现句 / 135
  - 6.3.4 三个时期翻译文本和原创文本的“是”字句 / 136
  - 6.3.5 三个时期翻译文本和原创文本的体标记 / 137
- 6.4 显化特征 / 138
  - 6.4.1 结构显化 / 138
  - 6.4.2 语义显化 / 144
- 6.5 多方验证：参考语料库的检验 / 147
  - 6.5.1 单语作家文本语料库 / 147
  - 6.5.2 清末文学文本语料库 / 149
  - 6.5.3 小结 / 150

## 第7章 翻译文本和原创文本历时变化比较的启示

---

152

- 7.1 翻译对于现代汉语变化的实证研究启示 / 152
  - 7.1.1 汉语译文的语体特征变化 / 152
  - 7.1.2 汉语译文的词汇特征变化 / 153
  - 7.1.3 汉语译文的句式特征变化 / 154
  - 7.1.4 汉语译文的显化特征变化 / 154
- 7.2 翻译作为语言变体对语言接触研究理论的启示 / 156
- 7.3 翻译文本特征的历时复合考察对于翻译共性的启示 / 159
- 7.4 影响因素的探析：语内因素和语外因素 / 160

第 8 章 结论	163
8.1 本书的突破及其对语料库翻译研究的贡献 / 165	
8.2 本书未来的研究方向 / 166	
附录 1 翻译文本三个时期高频词词表(前 100 个)	168
附录 2 翻译文本三个时期低频词词表(后 100 个)	172
参考文献	177
索引	195

# 图 目 录

- 图 2.1 翻译和语言接触研究文献变化(1984—2018) / 12
- 图 2.2 翻译和语言接触研究被引文献变化(2000—2019) / 12
- 图 2.3 2000—2019 年间每年发表的出版物数量柱状图 / 13
- 图 2.4 近十年语料库翻译研究领域前沿专题(2008—2018) / 14
- 图 2.5 翻译与语言接触研究中概念分析层级图 / 14
- 图 4.1 新型历时复合语料库的构成 / 40
- 图 5.1 维度 1 三个时期英语源语文本和翻译文本的历时变化 / 55
- 图 5.2 1930 时期和 1990 时期维度 3 英语源语文本和汉语翻译文本相关程度变化 / 58
- 图 5.3 维度 2(TCNAR)和维度 4(TCARU)三个时期英语源语和翻译文本的历时变化 / 59
- 图 5.4 翻译文本随机森林决策树预测图 / 62
- 图 5.5 汉语翻译文本三个时期重要性预测指标排序 / 62
- 图 5.6 三个时期英语文本和翻译文本词类分布聚类图 / 64
- 图 5.7 三个时期翻译文本高、中、低频词汇变化 / 65
- 图 5.8 三个时期翻译文本中四字成语的变化 / 73
- 图 5.9 三个时期翻译文本高频词搭配词目变化 / 75
- 图 5.10 翻译文本主语 it 在三个时期显化程度的变化 / 86
- 图 5.11 翻译文本宾语 it 三个时期显化程度的变化 / 88
- 图 5.12 三个翻译文本语料库 it 语义共现成分变化趋势 / 96
- 图 5.13 三个时期翻译文本词汇、句语义特征变化随机森林分布 / 103
- 图 6.1 汉语翻译文本和原创文本在维度 1 上的变化 / 109

- 图 6.2 汉语翻译文本和原创文本三个时期在维度 3 上的变化 / 112
- 图 6.3 汉语原创文本三个时期历时变化重要性维度预测图 / 115
- 图 6.4 维度 1 三个时期汉语翻译文本和原创文本译语透过性变化 / 118
- 图 6.5 维度 3 汉语翻译文本和原创文本译语透过性历时变化 / 122
- 图 6.6 维度 4 汉语翻译文本和原创文本译语透过性历时变化 / 124
- 图 6.7 三个时期汉语翻译文本和原创文本高频词和低频词频次变化 / 126
- 图 6.8 三个时期汉语翻译文本和原创文本话语重述标记的历时变化 / 131
- 图 6.9 汉语翻译文本和原创文本短语单位变化的对比 / 132
- 图 6.10 汉语翻译文本和原创文本三个时期习语频次变化 / 132
- 图 6.11 三个时期汉语翻译文本和原创文本“把”字句的使用频次变化 / 135
- 图 6.12 三个时期汉语翻译文本和原创文本存现句的使用频次变化 / 136
- 图 6.13 三个时期汉语翻译文本和原创文本体标记频次变化 / 137
- 图 6.14 类比语料库中三个时期指代衔接程度的对应分析 / 138
- 图 6.15 类比语料库中“它”字频率的变化 / 140
- 图 6.16 英语源语、汉语翻译文本和原创文本三个时期“它”字词汇的语义变化 / 142
- 图 6.17 1930 前后类比语料库 it 一词语义变化 / 143
- 图 6.18 1960 前后类比语料库 it 一词语义变化 / 143
- 图 6.19 1990 前后类比语料库 it 语义变化 / 143
- 图 6.20 三个汉语原创语料库“它是”关系语义变化 / 143
- 图 6.21 三个时期汉语翻译文本语义类别变化对应分析 / 146
- 图 6.22 “它”字在三类文本中的使用频率 / 149
- 图 6.23 “她”字在四类文本中的使用频率 / 149

# 表 目 录

- 表 4.1 语料库库容(DCC) / 41
- 表 5.1 三个时期英语源语与汉语翻译文本指标提取数据统计 / 52
- 表 5.2 1930 时期译文语料库前 100 词中部分超用关键词 / 66
- 表 5.3 1960 时期译文语料库关键词分析 / 67
- 表 5.4 1990 时期译文语料库关键词分析 / 67
- 表 5.5 三个时期译文语料库话语重述标记变化 / 69
- 表 5.6 翻译文本多字词簇三个时期变化 / 71
- 表 5.7 DCC 中最常用翻译文本 1930 时期三字词簇及排序 / 71
- 表 5.8 1930 时期译文语料库关键值最高的前 10 个动词搭配数目变化 / 74
- 表 5.9 三个时期翻译文本人称代词频次变化 / 80
- 表 5.10 翻译文本三个时期物称代词频次变化 / 81
- 表 5.11 翻译文本在三个时期指示代词频次的变化 / 81
- 表 5.12 翻译文本三个时期 it 和“它”频次的变化 / 83
- 表 5.13 it 和“它”充当不同语法成分的频次变化 / 84
- 表 5.14 it 的源语成分变化 / 84
- 表 5.15 翻译文本“它”字源语转换成分变化 / 91
- 表 5.16 it 共现成分的语义类别 / 95
- 表 5.17 翻译文本“它”的语义共现成分频次变化 / 95
- 表 5.18 they 和“它们”频次的比较 / 96
- 表 5.19 “它们”的源语转换变化 / 97
- 表 5.20 “它们”的源语对等成分变化 / 97
- 表 5.21 翻译文本三个时期语义类别分布 / 101

- 表 6.1 汉语翻译文本和原创文本特征指标数据统计 / 106
- 表 6.2 1960 时期维度 1 汉语翻译文本与原创文本之间回归关系检测 / 116
- 表 6.3 1930 和 1990 时期维度 1 汉语翻译文本和原创文本回归关系检测 / 117
- 表 6.4 维度 2 翻译文本各项指标与原创文本回归分析 / 119
- 表 6.5 维度 3 指称明晰性特征汉语翻译文本各项指标与原创文本回归分析 / 120
- 表 6.6 维度 4 汉语翻译文本各项指标与原创文本回归分析 / 122
- 表 6.7 1930 时期汉语原创文本库关键词 / 126
- 表 6.8 1960 时期汉语原创文本关键词 / 127
- 表 6.9 1990 时期汉语原创文本关键词 / 128
- 表 6.10 汉语原创文本三个时期联结程度的变化 / 129
- 表 6.11 三个时期汉语原创文本话语重述标记变化 / 130
- 表 6.12 类比语料库“它”字频次的变化 / 139
- 表 6.13 类比语料库中“它们”一词的历时变化 / 143
- 表 6.14 翻译文本指代显化特征对于汉语原创文本的透过性效应 / 144
- 表 6.15 三个时期汉语翻译文本和原创文本语义类别差异 / 145
- 表 6.16 单语作家文本、翻译文本和原创文本指代衔接程度相关性比较 / 148
- 表 6.17 三类文本中各类代词的使用频次 / 148
- 表 6.18 四个汉语原创语料库指代性显化变化 / 150