

外研社语料库研究系列

Translation and the Development of Modern Chinese
A Diachronic Multiple Corpora (1900-1949) Study

翻译与现代汉语白话文的发展

基于历时复合语料库（1900-1949）的研究



赵秋荣 著

Corpora

外语教学与研究出版社
FOREIGN LANGUAGE TEACHING AND RESEARCH PRESS

外研社语料库研究系列

国家社会科学基金青年项目 (12CYY001)

Translation and the Development of Modern Chinese
A Diachronic Multiple Corpora (1900-1949) Study

翻译与现代汉语白话文的发展

基于历时复合语料库 (1900-1949) 的研究



赵秋荣 著

图书在版编目 (CIP) 数据

翻译与现代汉语白话文的发展：基于历时复合语料库（1900—1949）的研究 / 赵秋荣著. — 北京：外语教学与研究出版社，2021.3（2021.11重印）

（外研社语料库研究系列）

ISBN 978-7-5213-2299-6

I. ①翻… II. ①赵… III. ①现代汉语—语料库—研究 IV. ①H109.4

中国版本图书馆 CIP 数据核字 (2021) 第 039788 号

出版人 徐建忠
项目负责 李晓雨
责任编辑 付分钗
责任校对 闫 璟
封面设计 黄 浩
出版发行 外语教学与研究出版社
社 址 北京市西三环北路 19 号 (100089)
网 址 <http://www.fltrp.com>
印 刷 北京盛通印刷股份有限公司
开 本 650×980 1/16
印 张 13
版 次 2021 年 4 月第 1 版 2021 年 11 月第 2 次印刷
书 号 ISBN 978-7-5213-2299-6
定 价 49.90 元

购书咨询：(010) 88819926 电子邮箱：club@fltrp.com

外研书店：<https://waiyants.tmall.com>

凡印刷、装订质量问题，请联系我社印制部

联系电话：(010) 61207896 电子邮箱：zhijian@fltrp.com

凡侵权、盗版书籍线索，请联系我社法律事务部

举报电话：(010) 88817519 电子邮箱：banquan@fltrp.com

物料号：322990001



记载人类文明
沟通世界文化
www.fltrp.com

前 言

“五四”新文化运动期间，汉语的词汇、句法和语篇结构等发生了显著变化，翻译在其中起到了至关重要的作用。为探究现代汉语白话文发展阶段中的变化及翻译在其中的作用，本书借鉴编码复制框架，设计出一个基于历时语料的复合语料库考察模式（1900-1949）：以文学历时语料为主，综合类比语料库、双语平行语料库、参考语料库与单双语作家语料库，将总考察期内的类比语料库分为四个研究阶段，以每十年的后五年为一个考察阶段。

主要研究内容包括：

（1）定量分析翻译汉语和原创汉语的语言文体特征。考察词汇密度、平均句长和平均句段长后发现：现代汉语词汇密度增加，句子长度增大，结构扩展。

（2）从词汇、句法和语篇视角选取典型案例，探究翻译对现代汉语发展变化的影响。词汇上，探究数量词的欧化，以“一个”为例。发现“一个”复制了译文的用法，频率增加、用法扩展。句法上，研究定语结构的欧化，以“一个+的+名词”中的定语结构为例。研究显示：现代汉语白话文的定语结构复制了译文表达，定语长度增加、结构扩展、容量扩增。语篇上，以话语重述标记为例，现代汉语白话文的话语重述标记复制了译文中的相关表达，频率增加，位置和功能逐渐固化。

（3）基于单、双语作家语料库考察了翻译对原创汉语的影响。欧化是“五四”作家的普遍现象，单语作家也不例外。在单、双语作家的共同作用下，

现代汉语白话文短时间内发生了欧化是此期白话文发展的一个显著特点。

(4) 探讨了翻译引起目标语言变化的机制。内因和外因共同作用决定了翻译影响目标语言变化的可能性、必然性和限度。

光阴似箭，从2010年开始攻读博士学位，至今已过十年。拙著即将完稿之际，回首这几年的求学历程，感慨万千。

能到北外学习，完成博士论文，走上翻译研究的道路，导师王克非教授是最重要的引路人。王老师言传身教，指引我阅读、参加学术会议、了解学术前沿、结识国内外学人、参与申请国家社科重大项目等，我从中受益匪浅。为了帮我迅速了解国内外最新进展，王老师总利用开会或其他机会把国内外最新文献带给我阅读，并鼓励我到国外学习。博士论文写作中更是层层把关，从选题、开题、撰写、修改，王老师都倾注了大量心血。这些年，师母周老师也特别关注我的成长，经常提醒我注意身体，关心我生活、学习、工作的点点滴滴。毕业之后，王老师和师母仍然非常关心我的研究、工作和家庭生活，给了我多方面的帮助和支持。

论文写作得到了多位学者的指点。开题时王逢鑫教授、张克定教授和陈国华教授提出了宝贵的意见和建议。感谢马会娟教授、梁茂成教授、封宗信教授、张政教授和刘泽权教授百忙之中参加我的论文答辩会，并给予宝贵的修改建议。

在北京外国语大学学习期间，刘润清教授的《语言哲学》、王克非教授的《双语语料库与翻译》、梁茂成教授的《语料库语言学》、许家金教授的《应用语言学中的计算机应用》、熊文新教授的《计算机辅助翻译》等课程开阔了我的视野。在英国曼彻斯特大学翻译与跨文化研究中心学习期间，Mona Baker 和 Meave Olohan 教授的《翻译研究方法》等课程对本书的撰写提供了帮助和支持。曼大的 Sofia 博士、汉堡大学的 Becher 博士、美国的 Kubler 教授对我研究中遇到的问题提供了诸多建议。在德国美因茨大学访学期间，Svenja Kranich 教授、Silvia Hansen-Schirra 教授、Lars Johanson 教授对我的研究方法和发现给予了热情的肯定和鼓励。在香港理工大学学习期间，李德超教授给予了诸多鼓励与帮助！

研究的顺利进行，离不开师兄、师姐们的帮助和支持。师兄秦洪武教授利用来京出差、讲学的机会多次指点迷津：分析工具和方法、语料考察、

研究途径、参考书目，事无巨细地帮助我。同门师兄张威、李京廉、黄立波、武光军、李越、杨志红、方红、黄焰结、刘瑾玉等为我的研究工作提供了很多建设性意见。书稿修改过程中我的研究生帮我做了很多校对工作，一并谢过！

家人的支持和帮助是我学业最有力的支撑。十年前，我爱人杜小平积极鼓励我到北京求学。他经常把他研究中的心得与我分享，鼓励我多进行跨学科研究。他虽工作繁忙，但兼顾了工作、家庭，还帮我开发软件、讨论研究中遇到的疑难问题。备战考博时，大女儿盼盼刚刚一岁半；书稿撰写完成时，二女儿欣欣已经四岁多。多年来，父母、公婆轮流帮我照顾孩子，解决了我的后顾之忧。十年来，寒暑假、节假日，挑灯晨读，夜深人静的冥思苦想，才有了今天这本书。她虽然非常稚嫩，但代表了我这段时间的学习经历和收获。

本书初稿写作中先后获得2010年北外创新计划、2011年北京市高等学校联合培养计划、2011年教育部人文社科基金、2012年国家社科基金等支持，评审老师们的厚爱给我提供了诸多机会，增强了我的信心，感谢这些基金支持我顺利完成拙著。

论文的部分章节已在第二届全国语料库翻译学研讨会、第四届基于语料库的语言对比与翻译研究(UCCTS4)、第八届国际语料库语言学(CILC)等会议上做口头报告，在剑桥学者出版社的*Studies in Language and Translation*、《中国翻译》《外语教学理论与实践》等期刊上发表，评阅老师们提出的宝贵建议为进一步完善本书提供了参考。

刘禹锡诗云，“莫羨三春桃与李，桂花成实向秋荣”，除了学术上的训练，十年来的学习和研究也是对自己品性和品行的锻炼。有了重压下处理学习、工作、家庭和生活等各方面的经历，将为今后研究工作打下坚实的基础。我相信，天道酬勤。这本书的完成是我学术生涯中重要的一环。今后，我会继续努力，争取取得更大成绩回报所有关心我、爱护我的师长、朋友和家人。

目 录

表格目录	VIII
插图目录	IX
第一章 绪论	1
1.1 研究缘起	1
1.2 研究假设及研究问题	3
1.3 研究设计	4
1.4 理论依据与研究方法	7
1.5 研究意义	8
1.6 本书结构	9
第二章 相关研究综述	11
2.1 翻译影响目标语言发展变化的研究	11
2.1.1 英化研究：基于相近语言对间的考察	12
2.1.2 欧化研究：现代汉语白话文的早期发展	20
2.2 语料库与语言变化研究	37
2.2.1 语料库分类	37
2.2.2 历史语料库与历时语料库	38
2.2.3 基于翻译驱动语料库的语言与翻译研究	41
2.3 小结	46
第三章 理论、方法与模式	48
3.1 编码复制框架理论	49
3.1.1 编码复制框架理论的定义	49
3.1.2 编码复制框架理论的类型	50
3.1.3 编码复制的影响因素	52
3.1.4 编码复制框架理论在翻译研究中的应用	55
3.2 语料库翻译学的理论与方法	57
3.3 基于历时复合语料库的考察模式	58
3.4 小结	61

第四章	历时复合语料库的建设与加工	62
4.1	语料库的总体架构	62
4.2	历时复合语料库的设计	62
4.2.1	语料选择的标准	62
4.2.2	历时复合语料库的语料选择	64
4.3	历时复合语料库:分段标准	69
4.4	语料收集、加工、标注、提取	70
4.4.1	语料收集	70
4.4.2	语料预处理	72
4.4.3	语料标注	73
4.4.4	语料检索、提取	78
4.5	小结	80
第五章	原创汉语与翻译汉语文本的语言文体特点: 基于统计的考察	81
5.1	考察语言文体特点的常用参数	81
5.2	翻译汉语和原创汉语的文体特征比较	85
5.2.1	词汇密度	85
5.2.2	平均句长	88
5.2.3	平均句段长	91
5.3	小结	93
第六章	翻译对现代汉语白话文发展变化的影响:基于语言层面的 微观考察	95
6.1	数量词的欧化	95
6.1.1	汉语原有结构与欧化结构的竞争与共存	97
6.1.2	“一个”用法的扩展	103
6.1.3	数量词欧化的译源考察	109
6.1.4	数量词欧化机制的探讨	113
6.2	定语封闭结构的欧化	114
6.2.1	定语长度	116
6.2.2	定语封闭结构的扩展	118
6.2.3	定语扩增的译源结构考察	120
6.2.4	定语封闭结构欧化机制考察	126

6.3	话语重述标记的欧化.....	129
6.3.1	使用频率.....	130
6.3.2	话语重述标记的分布.....	132
6.3.3	特定话语重述标记的句法和分布特点.....	134
6.3.4	总结类话语重述标记的译源考察.....	137
6.3.5	话语重述标记的欧化机制考察.....	139
6.4	单双语作家语料的比较.....	140
6.4.1	语言文体特征比较.....	141
6.4.2	语言微观层面对比.....	142
6.4.3	翻译对语言创作的影响.....	145
6.5	小结.....	148
第七章	翻译影响目标语言变化的机制.....	149
7.1	翻译影响目标语言发展变化的因素.....	149
7.1.1	语言因素.....	151
7.1.2	社会语境因素.....	153
7.2	翻译触发现代汉语白话文发展变化的机制.....	155
7.3	小结.....	157
第八章	结语.....	158
8.1	研究发现.....	158
8.2	研究贡献.....	161
8.3	研究局限.....	162
8.4	今后研究课题.....	163
	参考文献.....	164
	附录 1: 类比语料库 (1910-1949).....	183
1.1	原创汉语语料库.....	183
1.2	翻译汉语语料库.....	186
	附录 2: 双语平行语料库 (1930-1949).....	189
	附录 3: 参考语料库.....	191
	附录 4: 单双语作家语料库.....	192

表格目录

表 2-1	隐型翻译项目语料库	14
表 2-2	隐型翻译研究发现及研究成果	16
表 2-3	美式英语和英式英语对应语料库(采样间隔大约为 30 年)	40
表 4-1	类比语料库的构成及库容	70
表 5-1	翻译汉语和原创汉语各参数对比	85
表 6-1	“一个 + 名词”的使用频率(每万字的使用频率)	100
表 6-2	“个 + 名词”的使用频率(每万词的使用频率)	101
表 6-3	“一个”扩展形式的使用频率(百分比)	103
表 6-4	翻译汉语和原创汉语的平均定语长度比较	116
表 6-5	话语重述标记频率对比(每万字中的出现频率)	131
表 6-6	翻译汉语中话语重述标记的分布(百分比)	133
表 6-7	原创汉语话语重述标记的分布(百分比)	133
表 6-8	总结性话语重述标记的句法和分布特点(百分比)	134
表 6-9	话语重述标记“总之”“总而言之”的译源	137
表 6-10	单、双语、译者兼作家语言文体特征考察结果	141
表 6-11	“一个 + 名词”、定语结构和话语重述标记频率对比 (每万字的使用频率)	143
表 7-1	借用等级(Thomason and Kaufman 1988: 74-75)	150
表 7-2	英汉翻译影响现代汉语白话文发展变化的因素	156

插图目录

图 2-1	翻译驱动语料库的类型及应用	42
图 3-1	编码复制示意图	51
图 3-2	编码复制在翻译研究中的应用	57
图 4-1	语料库的总体架构	63
图 4-2	语料库加工路线图	71
图 4-3	文本处理工具	73
图 4-4	WordSmith 6.0 的主要功能	78
图 4-5	用 ParaConc 对齐语料	79
图 4-6	用 ParaConc 检索“总之”的译源结构	79
图 5-1	翻译汉语和原创汉语的标准类符 / 型符比	86
图 5-2	各阶段平均句长比较	88
图 5-3	各阶段平均句段长分布	92
图 6-1	“一个 + 名词”的提取	99
图 6-2	“一个 + 名词”的使用频率	100
图 6-3	翻译汉语和原创汉语定语平均长度比较	116
图 6-4	话语重述标记提取	130
图 6-5	话语重述标记使用频率对比	131

第一章 绪论

1.1 研究缘起

与万事万物一样，语言也处在不断发展变化中。影响语言发展变化的因素有很多，除了语言本身的变化即内因发挥作用外，外因也发挥了很大作用。外因主要指语言接触、语言政策和翻译政策以及各种合力的作用。翻译触发的语言接触和语言变化（translation-induced language contact and language change）在目标原创语言变迁中发挥了重要作用。

随着对翻译功能认识的不断深入，研究手段不断更新，尤其是大数据时代各种类型语料库的研发，翻译与目标原创语言发展变化关系的研究正逐渐成为国内外研究的热点和重点。国际上，相关研究主要集中在相近语言对之间的考察，尤其是英语翻译的影响。如英语翻译对德语（House 2006, 2011; Baumgarten and Özçetin 2008; Steiner 2008; Kranich et al. 2012; Kranich 2014; Bisiada 2018, 2019）、法语（McLaughlin 2011）、意大利语（Laviosa 2010）、丹麦语（Gottlieb 2004）、葡萄牙语（Hoey 2011）和希腊语（Malamatidou 2011, 2016, 2017, 2018）等的影响。

总体来看，这方面研究有如下短缺：（1）主要集中于相近语言对，对语言特征差异较大的英汉语言对之间的研究相对缺乏；（2）所据语料多为科技文本语料库，基于其他文本类型语料的考察相对欠缺；（3）研究方法上仅凭借类比语料库、平行语料库或综合二者但库容较小；（4）将翻译发生的相关社会背景纳入考察范围的研究还不充分。

“五四运动”前后短短几十年现代汉语白话文发生了显著变化。王力(1954/1984: 434)的考察发现:“从民国初年到现在,短短二十余年之间,文法的变迁比之从汉至清,有过之而无不及。文法的欧化,是语法史上的一桩大事。”现代汉语白话文短时期内发生了重大变化,而这种变化与“五四运动”前后翻译骤增是分不开的。从新文化运动开始,许多知识分子提倡借助翻译改造现代汉语白话文(傅斯年 1919b; 钱学同 1919; 郑振铎 1921; 周作人 1922; 鲁迅 1931; 胡适 1935等),身体力行积极进行创作实践与翻译实践。这些学者丰富的著译以及理论探索为现代汉语白话文的变迁提供了摹本及理论支撑。现代汉语白话文吸收了翻译中的语言表达形式,逐步朝“精密化”“明确化”的方向发展。具体表现在句子由简单变为复杂:一方面,把要说的话尽可能概括起来,成为一个完整的结构;另一方面,化零为整,把许多零星的小句结合为一个大句(王力 1980: 473-483; 史存直 2008: 358-381)。

使用频率(frequency of use)变化是语言变化的推动力之一(Steiner 2008: 336; Leech et al. 2009: 50),也是促成语法化与语言变化的核心(Hopper and Traugott 2003: 129),而基于语料库的数据描述、统计、分析是观察频率变化的有效途径。历时语料库为准确追踪语言变化提供了前所未有的条件(Leech et al. 2009: 28)。

在积极借助翻译改造现代汉语白话文的特殊社会语境下,在与外语的接触过程中,翻译汉语模仿和复制了源语的结构与形式。或者说一段时期内源语渗透(shining-through)(Teich, 2003)发挥了主导作用,现代汉语白话文模仿了翻译语言的表达形式,在社会环境、语言发展等多重因素影响下,最终促成了现代汉语白话文短时间内的变化。

现代汉语欧化研究一直是学界关注的重点。较早的研究有王力(1943/1985, 1954/1984, 1980)、北京师范学院(1959);海外或港台研究有Tsao(1978)、Kubler(1985)、Gunn(1991)、Hsu(1994)、谢耀基(1990)、石定栩、朱志瑜(1999, 2000),近期研究有王克非(2002)、刁晏斌(2006a, 2006b)、贺阳(2008)、朱一凡(2011)、王克非、秦洪武(2009, 2013)、戴光荣(2013)、赵秋荣、王克非(2013, 2014, 2020)、秦洪武、夏云(2017)等,这些研究为本书提供了重要基础。但已有研究或依靠主观经验判断,或基于典型但有限的实例,或集中于当代语料,或侧重描述语言

现象，理论探索相对欠缺。定量考察现代汉语白话文的历时变化，从语言接触视角揭示翻译在现代汉语白话文发展变迁中的作用、探讨翻译引起目标原创语言变化的特点、规律及机制的研究仍然比较缺乏。

综上所述，翻译影响目标原创语言变迁的研究已经引起了众多研究者的关注。此类研究不仅语料上需要进一步扩充，研究方法上也需要进一步探讨，如综合不同类型语料库的优势。不仅需要共时研究，更需要历时视角的观察，还需要将翻译发挥作用的社会语境纳入考察范围，以及从理论上探讨翻译引发目标语言发展变化的机制。因此，基于大规模历时复合语料库的考察模式将是目前较有效的尝试。该方法不仅综合了各类语料库的长处，还将为定量描述现代汉语白话文的发展变化、揭示翻译的社会作用提供有力支持，也有利于进一步揭示接触视角下语言发展变化的规律。

1.2 研究假设及研究问题

基于考察现代汉语白话文的历时变迁及翻译的作用，本书的假设是：

(1) 现代汉语白话文逐步走向精密化、明确化，这一过程与翻译有密切关系。语言自身的发展一般较为缓慢，如果某种语言较短时间内发生剧烈变化，多来自外因的影响。“五四”前后较短时间内现代汉语白话文发生了重大变化，当时社会上许多具有影响力的知识分子提倡借助翻译改造现代汉语白话文。可以假设翻译中高频出现的、欧化的或新的表达方式在这些作家作品中的出现频率可能更高。

(2) 社会语境在现代汉语白话文早期发展中起了关键作用。“五四”新文化运动为现代汉语白话文吸收翻译汉语语言表达、促进自身发展提供了社会基础。欧化表达方式在当时的书面语尤其是译文中有多体现，而创作语言中欧化的程度却有所不同：受源语渗透效应影响，双语作家可能直接吸收外语中的语言表达；单语作家可能模仿他人译文中的新表达方式。但只有单、双语作家共同作用才可能推进现代汉语白话文短时间内的变迁。因此，社会语境发挥的作用不容小觑。

基于此，本书思考三大问题：

(1) 现代汉语白话文向精密化、明确化方向发展的过程中，词汇、句法和语篇方面发生了哪些变化？

(2) 翻译在现代汉语白话文早期发展过程中是否发挥了作用？如果是，现代汉语白话文模仿、借鉴或吸收了翻译汉语的哪些表述形式？单、双语作家吸收外来语言的表现是否有差异？

(3) 语言接触视角下翻译触发目标语言变迁的特点是什么？有何规律可循？

1.3 研究设计

王力（1980：14-17）指出：研究现代汉语白话文的历时变化应该把握一些原则，如注意语言发展的历史过程；密切联系语言发展的历史；重视语言方面的联系以及辨认语言发展的方向。向熹（1993：6-8）归纳现代汉语的研究方法为：归纳、比较、统计、实证、探源、转换、推演等。Kubler（1985：146）则提出建设语料库，定量观察语言变化。他尤其指出：对比考察欧化前后的语料将是研究现代汉语白话文历时变化的理想方法。

基于前人研究语言变化的经验与研究成果，本书认为：考察现代汉语白话文的历时变迁需要以历时语料为基础，加密时间间距，分段考察，以凸显语言的历时变化；现代汉语白话文的历时变化在语言文体特点、词汇、语法和语篇等多个层面都可能有所体现。此外，现代汉语白话文短时期内发生了重要变化，此变化既是可能的，又有一定限度，需要结合具体社会语境讨论其变化的可能性与限度，拓展翻译触发目标语言变化的机制。

鉴于翻译与目标语言相互关系的复杂性和多种类型语料库的欠缺，本书借鉴编码复制框架理论（code-copying framework）和语料库翻译学相关理论，设计出一个基于历时语料的复合语料库考察模式，从词汇、句法和语篇视角选取典型案例，探究翻译与现代汉语白话文早期发展之间的关系与影响。

现代汉语白话文发展初期即五四前后，文学翻译是大规模翻译活动的主体，因此，本研究将主体语料锁定为小说、散文、戏剧（以及少量社会科学文本）等体裁的书面语语料。历时复合语料库共包含四个语料库：历时类比语料库（1910-1949）、双语平行语料库（1930-1949）、参考语料库和单双语作家语料库。具体设计如下：

(1) 历时类比语料库。旨在考察现代汉语白话文早期的历时发展, 比较原创汉语和翻译汉语语言特征的差异。历时类比语料库主要包括原创汉语和翻译汉语两个子库, 语料来源集中于1910年到1949年间的语料。原创汉语语料包括鲁迅的《狂人日记》、周作人的《雨天的书》等, 库容为1,193,529字; 翻译汉语选取在中国翻译史上产生深远影响或广为人知的译作, 如《傲慢与偏见》《鲁滨逊漂流记》等, 库容为960,804字。类比语料库的总库容为2,154,333字。

分段历时考察语言变化是定量直观观察语言变化的常用方法。国际上诸多大型语料库多采用历时分段的方法。如赫尔辛基语料库以100年为一个阶段观察古英语到早期近代英语的变化; 布朗家族语料库以30年为一个分期分析当代英语语法的历时变迁; 隐性翻译语料库 (Covert Translation) 以20年为一个分期探索翻译对德语的影响。国内历时变迁研究多以历史事件为划分标准。“五四”前后现代汉语白话文的发展有一定特殊性, 尤其是短短二三十年间语言发生了重大变化, 如果分段时间过长, 有些变化可能很难捕捉到; 如果以自然年为分段标准, 则很难确保语料能覆盖所有时间段; 如果以历史事件为分段标准, 历史事件与语言发展的直接关系也不确定。因此, 无论长时间段划分法、自然年划分法还是以历史事件为基础的划分法, 都无法直接运用于考察翻译与现代汉语白话文早期变化研究中。基于上述讨论, 本书从现代汉语白话文早期发展的实际情况出发, 以自然年每十年的后五年为一个考察阶段, 合理划分时间间隔并加密考察间距。具体为: 1915-1920为第一阶段, 1925-1930为第二阶段, 1935-1940为第三阶段, 1945-1949为第四阶段。

(2) 双语平行语料库。旨在考察翻译汉语的译源结构, 也就是通过查找译本结构中对应的源语语言结构表达方式, 揭示现代汉语白话文早期发展过程中源语 (或翻译) 的作用。受时代、语言质量及可获得的英汉双语语料的限制, 本书中双语平行语料库的译语为现代汉语白话文 (不含文言文), 且尽可能涵盖多位译者。

参考挪威奥斯陆大学的英挪双语平行语料库 (The English-Norwegian Parallel Corpora, ENPC) 和北京外国语大学“超大型英汉双语平行语料库”的抽样标准, 本书采用全文录入与部分抽样相结合的方法: 短小语篇全文

录入，较长语篇提取作品开头的连贯语句，取样多在1.5万到3万字/词之间。时间跨度在1930-1949年间，库容为1,463,534字/词。

(3) 参考语料库。参考语料库主要选取未受欧化影响的旧白话作品。由于类比语料库为受欧化影响的语料，对比参考语料库与类比语料库中的语料，定量观察现代汉语白话文欧化前后的语言特点及变化，可以确定“五四”后哪些语法现象是在翻译影响下新出现的，哪些是扩展了原有用法，哪些是激活了原有用法逐渐发展起来的。王力（1943/1985，1954/1984，1980）研究现代汉语白话文欧化语法现象时，主要参考语料为《红楼梦》和《儿女英雄传》。王力（1980：472）指出：通过对比现代书报和参照语料可以发现欧化的语言现象，但缺点是语料量相对较少，只能是举例式的。Kubler（1985：146）在研究中也明确指出：参考语料库在研究现代汉语白话文变迁中具有不可忽视的作用。基于此，本书参考语料库主要选取未受欧化影响的中国语的典型代表——清代旧白话小说作品，如《红楼梦》《儿女英雄传》等，参考语料库的库容为4,071,420字。

(4) 单双语作家语料库。旨在对比双语作家与单语作家创作语言中语言特征的异同，佐证翻译语言对现代汉语白话文创作语言的影响。“五四”新文化运动前期，许多倡导者认识到萌芽期的现代汉语白话文不能表达精密的思想，而翻译可以作为改造现代汉语白话文的“利器”。王力（1954/1984：349）提醒我们：“欧化的来源是翻译，译品最容易欧化。欧化语法的来源是译品、准译品和以西语为腹稿的作品”。“五四”时的大多数知识分子有国外学习的经历，熟练掌握一门或多门外语，身兼“作家”与“译者”双重职责，著译颇丰。这些先行者既能创作又能翻译，且欲借翻译改造现代汉语白话文，可以设想，他们在翻译过程中会积极吸收西方语言的结构形式；创作中大量模仿翻译中高频出现的、新的语言形式。“五四”时期，欧化是文学领域的普遍现象。欧化的语言特征不仅出现在双语作家、译者兼作家的作品中，还出现在单语作家的作品中。如果说双语作家的欧化语言受到外语和翻译的双重影响，单语作家的欧化语言则只能受到翻译语言的影响。单双语作家语料库的考察将有利于揭示翻译语言对汉语创作语言的影响，探讨现代汉语白话文发展初期如何模仿翻译语言促进自身发