

语料库翻译学文库

国家社会科学基金项目资助（项目批准号：10BYY008）

基于历时语料的 翻译与现代汉语 互动研究

秦洪武 夏云◎著



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

语料库翻译学文库

国家社会科学基金项目资助（项目批准号：10BYY008）

基于历时语料的 翻译与现代汉语 互动研究

秦洪武 夏云◎著



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

内容提要

本书采用新方法探讨翻译对现代汉语的影响以及现代汉语对翻译中语言运用的制约作用。

研究发现,现代汉语在历时发展过程中最明显的变化表现在结构容量、结构构成和话语组织方式上,这些都与同期翻译语言的特征相似;汉语翻译语言曾经大幅偏离汉语的常规特征,主要表现在扩展结构容量、多用话语标记,这些特征在现代汉语中存留下来,已成为汉语使用的主要特征。从另一角度看,汉语翻译语言受汉语常规特征的约束,但未能充分挖掘汉语固有的丰富的表达资源。总之,语料库为翻译和目标语之间的互动研究提供了更宽广的视野和更有力的方法论支持,也将使系统解析语言演化机制成为可能。

图书在版编目(CIP)数据

基于历时语料的翻译与现代汉语互动研究 / 秦洪武, 夏云著.

—上海: 上海交通大学出版社, 2017

(语料库翻译学文库)

ISBN 978 - 7 - 313 - 17705 - 6

I. ①基… II. ①秦… ②夏… III. ①汉语-翻译-研究

IV. ①H159

中国版本图书馆 CIP 数据核字(2017)第 177863 号

基于历时语料的翻译与现代汉语互动研究

著 者: 秦洪武 夏 云

出版发行: 上海交通大学出版社

邮政编码: 200030

出 版 人: 谈 毅

印 刷: 常熟市文化印刷有限公司

开 本: 787mm×960mm 1/16

字 数: 353 千字

版 次: 2017 年 9 月第 1 版

书 号: ISBN 978 - 7 - 313 - 17705 - 6/H

定 价: 78.00 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: 19

印 次: 2017 年 9 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: 0512 - 52219025

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”、“基于语料库的莎士比亚戏剧汉译研究”、“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

前 言

语言变动不居,语言之间可能相互影响。语言变化是语言研究永恒的话题,从古代“小学”、语文学到历史语言学,再到语言类型学,语言变化似乎从未游离人们的视野。及至目下,在信息技术时代,语言处理技术日臻成熟,语言演化的更多层面和细节业已纳入观察范围,愈加容易捕捉、提取和归类。不仅如此,得益于信息技术,语言演化的研究方法在不断丰富、优化和完善,我们可在更广范围内、更多层面上观察语言构成的历时变化特征。本书中的研究就是在此背景下展开的。

本书使用新方法探讨翻译对现代汉语的影响,以及现代汉语对翻译中语言运用的制约作用。在 20 世纪初期,翻译对现代汉语产生过影响,这是老话题,但有许多问题到目前也未能有明确的回答。比如,“欧化”是否真的发生?如果有,为什么没有新的语法规则产生?如果没有,为什么 20 世纪 20 年代的汉语会明显不同于之前的汉语白话?如果现代汉语的变化不是单一语法范畴的增损,不是汉语语法规则的变化,也不是基本语序的变化,那么,汉语的演化很可能来自其内部构成,即汉语结构的内部在其构成上发生了变化,进而导致表层形式发生变化。本书尝试基于该假定考察汉语翻译语言对现代汉语可能造成的影响和影响方式。

传统上,有关汉语翻译对现代汉语发展影响的研究大多注重观察汉语中某类或某一项目使用上的变化,极少注意与此相关的其他语言项目的变化,这样就难以捕捉语言的系统演化,更无法对汉语发展过程中的翻译因素做出合理的解释。鉴于此,本书基于语料库,以频率为依据,通过对比和类比分析观察汉语发展及汉语翻译语言在较大时间跨度内的变化和发展状况。在实施比较和分析时,我们尽量不从现有的文献中预设研究的语言项目和现象,而是直接通过对比数据发现汉语翻译和原创语言之间可能存在的差异,以及汉语本身可能存在的历时差异。这样做既可以减少语言描述的主观性,又可以发现许多传统方法难以发现的变化特征。期待这一方法能为汉语研究和历史语言研究提供有益的分析思路。

本书发现,汉语翻译语言对现代汉语的影响,除了学界广泛认可的词汇、表达形式等层面上的影响外,突出表现在以下几个方面:①汉语在历时发展过程中最明显的变化表现在结构容量、结构构成和话语组织方式上,这些都与同期翻译语言的特征相似。具体说来,汉语结构容量较旧白话出现了明显的扩增,最终导致句段偏长。汉语句子较长的说法不严谨,其实质是句段偏长;汉语封闭结构的内部构成有

了重要变化,突出表现在它容纳数量更多、句法更复杂或者非平衡的表达形式,导致复杂定语出现;汉语话语组织方式的变化突出表现在新型话语标记的使用和使用频率,程序性和情态性话语意义的表达开始倾向于显性表达。②汉语翻译语言在 20 世纪 20 年代至 40 年代大幅偏离汉语的常规特征,但在 50 年代后开始靠近常规,主要表现为控制容量,但多用话语标记的语言使用特征却沿传下来。③汉语翻译语言受汉语常规特征的约束,但汉语翻译语言未能充分挖掘汉语固有的丰富的表达资源,这在叠用和助词的使用上表现较为明显。

简言之,现代汉语有别于汉语白话,其中的一些差异可能源于汉语翻译语言的影响。但需要指出,这些影响不是汉语词法规则的变化,也不是汉语句法的变化,更不是对外语尤其是英语语法规则的借用,而是对汉语表达资源的挖掘或者过度使用,故不应简单地归为“欧化”。

从事学术研究,苦与乐相伴,喜悦和迷茫并随,引以为豪的是我们未放弃对真知的追寻。感谢国家社科基金的资助和支持,感谢曲阜师范大学社科处和外语学院对本项目的鼎力相助。在研究过程中,有志趣相投的研究者如王克非教授、黄万丽老师和孔蕾博士等的支持、鼓励和帮助,有张丹、王玉、司佳冰、刘丹丹等同学的时间、精力和智力上的巨大投入,纵使荆棘载途,拙作亦能在最后期限前著就付梓。在此,我们要对为本书倾力奉献的参与者致以深深的谢意。

秦洪武 夏云

目 录

引言	1
----------	---

第一部分 类比语料库的创建和检索

第 1 章 历时语料库和历时翻译语料库的创建:理论和方法	6
1.1 研制汉语历时类比语料库的意义	6
1.2 历时语料库的创建和研究	7
1.3 历时语料库:构成、检索和研究	12
1.4 汉语历史语料库	17
1.5 小结	18
第 2 章 英汉平行历时语料库的建设和检索设计	20
2.1 平行语料库的建设状况	20
2.2 英汉平行语料库的创建	21
2.3 语料库检索平台的设计	28
2.4 检索系统支持的历时平行语料检索方式	28
2.5 小结	33
第 3 章 汉语语料检索中正则表达式的应用	35
3.1 非正则表达式的语料检索	35
3.2 使用正则表达式的语料检索	36
3.3 正则表达式用于汉语词、词组检索	38
3.4 正则表达式用于汉语检索:句段、结构	40
3.5 使用正则表达式检索应注意的问题	44

第二部分 基于类比语料的翻译与现代汉语互动研究

第 4 章 汉语历时演化研究概览	48
4.1 研究现状	48
4.2 语言演化研究方法	49
4.3 现代汉语演化研究:主要内容	52

4.4	现代汉语语言演化研究:主要发现	54
4.5	小结	57
第5章	汉语翻译语言与现代汉语的发展研究:内容、思路和语料	58
5.1	主要内容	58
5.2	研究思路和方法	58
5.3	语料来源	60
第6章	汉语翻译语言与汉语的互动:汉语翻译语言的描写性特征	62
6.1	基于语料库的翻译语言研究	62
6.2	汉语翻译语言的描述和分析方法	69
6.3	宏观描述	70
6.4	词性分布	71
6.5	组合方式	75
第7章	汉语原创与翻译语言的互动:历时比较	78
7.1	类比语料库	79
7.2	现代汉语发展和现代汉语翻译语言发展变化的类比分析	81
第8章	汉语翻译语言的微观描述:叠用词、结构容量、句段长度和 “被”字句	122
8.1	汉语翻译对汉语表达手段的调用:叠用词	122
8.2	翻译与现代汉语结构容量的变化	131
8.3	“被”字句语义特征和结构容量的演化	141
第9章	翻译与目标语发展的互动:话语组织方式的历时变化	151
9.1	汉语中的句首话语标记	152
9.2	句首话语标记的句法位置	152
9.3	句首话语标记的功能	153
9.4	研究方法和类比语料数据	154
9.5	言据类句首话语标记	156
9.6	详述类话语标记	166
9.7	推理类话语标记	175
9.8	对比类话语标记	182

第三部分 类比研究发现在翻译研究中的应用

第10章	多译本分析的结构容量视角	196
10.1	引言	196
10.2	句段内的结构容量	196

10.3	研究数据来源:英汉多译本翻译语料库	197
10.4	句长及句段长	198
10.5	结构式容量	201
10.6	对结构容量的解释	206
第 11 章	多译本对比和评估:基于语料库的重译研究:Robinson Crusoe 的	
	两个中译本	208
11.1	研究内容:Robinson Crusoe 的两个汉译本	209
11.2	常规语言特征描述	209
11.3	微观语言特征分析	212
第 12 章	结束语	220

附录

附表 1	1919—1930 年原创 vs 1911 年前原创词性框架 Keyness 分析 (两端绝对值最高前 30 例)	222
附表 2	1919—1930 年汉语翻译 vs 1911 年前汉语原创词性框架 Keyness 分析(2-5 POS 符)	225
附表 3	1911 年前、1919—1930 年代汉语原创及汉语翻译语言词性框架 可能关联的数值	243
附表 4	汉语推理类话语标记使用的历时变化	262
附表 5	汉语详述类句首话语标记的分类及历时变化(括号内数字为出现 频次)	264
附表 6	汉语详述类句首话语标记的历时变化(新出现的表达形式)	266
附表 7	用于汉语检索的正则表达式常用元字符	268
参考文献		270
索引		285

引言

语言在接触状态下会相互影响,但预测这种影响是十分艰巨的工作,有时甚至无法预测。根据 Siemund,自 20 世纪 50 年代以来,语言接触领域的大量研究大大加深了人们对语言接触的认识,但语言接触涉及的原则和机制,以及它们之间的互动关系依然难以把握(Siemund,2008:1)。本课题探讨的语言接触较为特殊,它是间接性的语言接触,即汉语翻译语言与汉语之间的互动和相互影响,这种影响的广度和深度远不及直接语言接触,支撑这种互动的原则和机制目前尚难以觅寻。就现代汉语而言,在其发展初期的确经历了一次重要的变化过程,该过程早已引起学界关注(王力,1943),并在当代取得了令人瞩目的进展(王克非,2002;贺阳,2008;朱一凡,2011;崔山佳,2013)。但迄今为止我们还没能对这种突出、突现的变化进行相对宏观的、系统的观察和描述。而且,随着信息处理技术的发展,有理由期望拓展观察范围,并相应提高研究深度,丰富研究内容。本课题就是在这一背景下实施研究的,我们祈望我们所做的努力能为丰富和充实该领域研究做出贡献。

本项研究是在国家社科基金“历时类比语料中的翻译与现代汉语发展研究”的结题报告基础上写成。按照本项目原初的研究计划,本课题研究包含三个阶段:建库阶段、研究阶段、语料库开放应用阶段。这三个阶段均已完成,但在研究过程中,我们根据资料收集的难度和工作进度对各阶段的工作进行了调整,实际完成情况如下:

最终建成的语料库还是由两个分库构成,即英汉翻译历时平行语料库和汉语原创可比文本库。但分库内部的构成有变化。

首先,原计划建设“现代白话文之前的翻译:1850—1914”子库,但由于这一时期的翻译(即便使用白话)在很大程度上属于改写,删减添增过于随意,研究价值不高,故没有收取该期的翻译文本,也未单独设置子库。

其次,“英汉翻译历时平行语料子库”中增加了“大众语运动之后的翻译:1935—1949”,目的是和前一时期的翻译分开,观察大众语运动之后对翻译的影响。

再次,新建成的库中,汉语可同时采用词和字的计数方式,以下是以字为单位计数的库容。

英汉翻译历时平行语料分库(1 305.7 万字/词)

现代白话文走向成熟时期的翻译:1918—1930(114.3万字,76万词,英汉双语)

大众语运动之后的翻译:1935—1949(115.7万字,76.5万词,英汉双语)

新中国成立初期的翻译:1950—1966(54.2万字,43.3万词,英汉汉语)

1980年代至今的翻译:1980年至现在(458.7万字,367万词,英汉双语)

汉语原创可比文本分库(370万词):

现代白话文之前的汉语原创文本:1911年前(122.9万词,汉语)

现代白话文走向成熟时期的原创文本:1918—1930(123.6万词,汉语)

新中国建立初期的汉语原创文本:1950—1966(40万词,汉语)

1980年代至今的汉语原创文本:1980年至现在(83.4万词,汉语)

1. 历时翻译库的库容

在实际建库过程中,我们考虑到样本过小,不易发现一些微观语言使用问题,因此大幅扩展了库容(2 000余万词),英汉翻译历时平行语料分库库容扩展了10余倍,汉语原创可比文本分库也比原计划大5倍。采样比原计划更多、库容更大、时间分布更趋合理。在此基础上实施的统计分析也更可信、可靠。

2. 历时翻译库的抽样框架

按照原先的设计,鉴于早期文本收集困难,需要分时段选取样本;同时,考虑语言之外的社会政治因素,将历时语料的抽样分成三个时段:时段一,以“五四”前后为分水岭,之前的仿旧白话翻译(黎锦熙,2000)和汉语原创白话作品作为参考库;时段二,自“五四”之后至1930年现代白话文经历巨变而走向成熟的时期(张中行,2007);时段三,从1950年代到1966年,时段四,自1980年到现在。但在实际建库时,我们根据研究需要并充分考虑了语料分段采集可能引发的问题,调整了原来的计划。即不再按时段选取样本,而是根据样本出版的年份不间隔时段来收录样本。这样可以避免人为干预样本的选取,研究时根据需要任意确定时间段来抽取样本,这样应更有利于揭示翻译和汉语本身的互动历程。

在建设语料库过程中,由于新闻类文本不易找到英语源语文本,故原来设想的新闻文体所占比例无法实现,但哲社类文本达到35%以上。

3. 语料处理

语料库力求科学建置(Sinclair, 2004)。在合理抽取样本的基础上,双语语料使用Aligner软件实施句对齐,使用CLAWS和ICTCLAS对英汉语语料进行POS标注和分词,并对部分语言和翻译现象进行人工标注。除使用常用的语料检索工具(Concordancer、Paraconc、WordSmith 5、MLCT等)外,还要充分利用文本类型、文体、时间等标头元信息,开发出针对性强的历时语料检索工具,该工具已经开放了在线检索。

4. 研究步骤

本课题按照预先计划的先描述后分析的思路着手研究。首先,将类比库中四个平行阶段在词汇(语素)、组合、构句以及组篇上的异同和变化趋势描述出来。其次,对描述和对比分析的结果进行分析。

5. 语料库开放应用阶段

按照研究计划,语料库建成后要实施多层面的立体、历时分析以揭示翻译语言与目的语之间的互动关系,并为这一互动关系建立历时的变化模式。通过宏观和微观层面的语言变化分析,探讨封闭结构容量变化和汉语篇章组织方式变化与翻译语言之间的关系,借此描述翻译和现代汉语变化之间的互动关系。

本研究克服了 20 世纪上半叶以前的语料收集的各种困难,并解决了复合式检索程序的开发问题,包括正则表达式的研究和封闭结构提取程序等。研究中的难点问题在课题研究过程中得到了很好的解决。

本研究分为三个部分。

第一部分探讨与类比语料库创建相关的问题。第一章探讨历时语料库的创建理论、原则和方法,重点探讨历时语料的抽样结构和检索方法,以及目前业已取得的研究成果。第二章集中探讨英汉平行历时语料库的创建和检索问题。有针对性地提出了 20 世纪英汉翻译历时语料的取样、对齐、标注和存储方法,重点探讨英汉历时平行语料检索平台的设计和检索问题。第三章聚焦于汉语语料的检索和提取方法,为此后的历时语言研究提供可靠的技术保证。

第二部分基于类比语料探讨翻译与现代汉语之间的互动。该部分包含六章。第四章总述和总结了汉语演化研究的内容、方法和演化模式,并指出汉语历时研究中存在的问题。第五章关注汉语演化过程中的翻译角色,从理论上规划了研究内容,以量化为特征的研究思路以及研究需要的类比语料。第六章基于双语语料库,从宏观层面描述了汉语翻译语言的描写性特征,涉及句长、句段长、结构容量、形类符比、词性分布等等,这也为它与汉语原创的类比奠定了基础。第七章从语块、词性块、封闭结构的构成角度,基于历时类比语料,探讨原创汉语演化过程与翻译语言历时变化可能存在的平行关系。第八章聚焦于叠用词、结构容量和句段长度等体现汉语特色的语言特征,观察翻译在现代汉语发展初期所起的重要作用。第九章则把目标转向现代汉语话语组织方式的重要变化,观察现代汉语话语标记演化过程中翻译语言可能扮演的角色,重点探讨言据类、详述类、推理类和对比类话语标记。

第三部分关注汉语翻译与汉语演化互动研究对汉语翻译语言研究的启示。第十章谈结构容量在多译本分析中的使用,第十一章使用语料库进行多层面、多角度分析重译。研究表明,类比语料分析不仅有助于观察汉语演化中的翻译因素,还有助于充分描述和评估汉语翻译语言。



第一部分 类比语料库的创建和检索

第1章

历时语料库和历时翻译语料库的创建： 理论和方法

汉语是字本位语言,缺少形态手段,词性和语法属性缺少形式依据;汉语更是少用形式手段表达句间逻辑关系。这些都给语料对齐、分词和检索带来困难,历时语料库的标注、检索以及类比研究也需解决上述问题,需要找到有效的应对方法。历时语料中各年代语言样本可以相互比较,发现语言运用的变迁过程,能为翻译研究、语言演化研究,以及翻译语言与目的语之间的互动提供数据支持。本研究认为,对历时语料的复合检索能够实现按时间、文类和其他文本属性信息检索语言数据;正则表达式检索还可弥补汉语分词和标注精确度上的不足,提高检索质量。此外,篇头数据和多样的数据检索工具可以在技术上保证对汉语语料进行多层面的比较和分析,从而有力地支持汉语历时研究。本章探讨历时语料库的研制目的、研制思路,并报告在这些方面业已取得的进展。

1.1 研制汉语历时类比语料库的意义

语言变化有多种研究路径,如语言生态学、类型学、社会学和语言接触(Hurford et al., 1998; Hickey, 2003)。语言接触涉及外来语言对特定语言的影响,是语言变化研究中的重要领域。语言接触分为直接语言接触和间接语言接触。前者的典型环境是双语社会,后者的典型环境是翻译。已有研究表明,这两类接触都属于双语交际,而这又有明显差异。直接语言接触会产生新的接触语言变体(contact language varieties),该变体为某一语言社团使用;而间接语言接触也就是翻译并不能产生新的语言变体,它只是翻译语言(translated language),也不是特定语言社团使用的语言变体。(Lanstyák & Heltai, 2012)。

按照 Lanstyák & Heltai 的观点,接触语言变体用来描述(descriptive),即日常交流信息,而翻译语言用于释义(interpretive),即只在翻译时使用。这样看来,翻译语言很难影响到目标语。然而,情况并不如此简单,根据 Shi(2006)的研究,翻译语言完全可能引发语言变化,尤其是书面语变化。如在香港使用的粤语中就出现语言结构的借用,这种借用来自英汉翻译,是这种低质量的直译将这些结构带入香港书面语。

总的看来,在有关语言接触和变化的文献里,间接语言接触很少受到关注,翻译自然也就不是关键词。而实际情况是,在特定历史时期,翻译对目标语有重要影响。比如,在文艺复兴时期,德国宗教领袖马丁·路德通过《圣经》翻译促成德语新文体产生;法国的安岳借助翻译改良、丰富了法语(王克非,1997)。在中国,近代汉语受佛经翻译影响而演化出了白话新文体(梁晓虹,1992);而且,在20世纪,现代汉语白话文还在翻译语言的促动下而走向成熟,同时它又制约着翻译语言的变化范围。二者如何相互影响?背后的动因是什么?这些问题需要通过语料梳理做出回答。

根据王力所言(1989),“五四”以后,汉语受西洋语法影响,经历了巨大变化,主要表现在特定词性和句式的使用频率、句法形式严密化,以及分句位置的变化上。这方面的研究关注到了语言的多个层面,涉及词汇、句法和语序,并把英语源语纳入观察范围(贺阳,2008)。长期以来,对汉语欧化的判断主要依赖研究者的自然语感(朱一凡,2011),研究的素材来源不够丰富,考察范围有限,研究的深度和广度明显受到制约;另外,汉语自身对汉语翻译语言的制约作用并没有引起足够重视,二者间存在的动态互动关系需要借助历时语料数据来观测和分析。

1.2 历时语料库的创建和研究

历史语言研究基于过去的语言素材,天然亲近经验性语言数据;这一传统沿传至今,并在当代有了新的发展。Lüdeling & Kytö(2008: vi)指出,现代语料库语言学得益于历时比较语言学,因为后者使用的研究素材总是文本或者文本的集合。而且,19世纪发展起来的语言谱系识别技术一直沿用至今,语言变化研究依旧仰赖早期的文本(也就是语料)。在此背景下,在欧美国家,自20世纪80年代以来,随着人们对语言历史发展兴趣重燃;随着计算机和信息技术在语言处理领域的广泛应用,可机读的便于检索和提取语言数据的历时语料库得到了飞速发展。基于这类语料库的语言研究蔚然大观,研究方法日益成熟,卓见迭出。在中国,历时语料库的创建和基于历时语料的研究起步较晚,语料库设计思路和研究思路较西方还远不够成熟,迫切需要对国外历时语料库的创建和研究有较多的了解。同时,基于语料库的历时语言研究本身也出现了许多问题和困难,这些都需要我们认真审视和总结。本章回顾二十多年来历史(时)语料库的发展状况,重点探讨该类语料库的设计、制作,以及基于这类语料库的语言研究。

1.2.1 历史语料库和历时语料库

研究语言变化,必然要用过去的语言素材。是过去的,自然就是历史的。但历