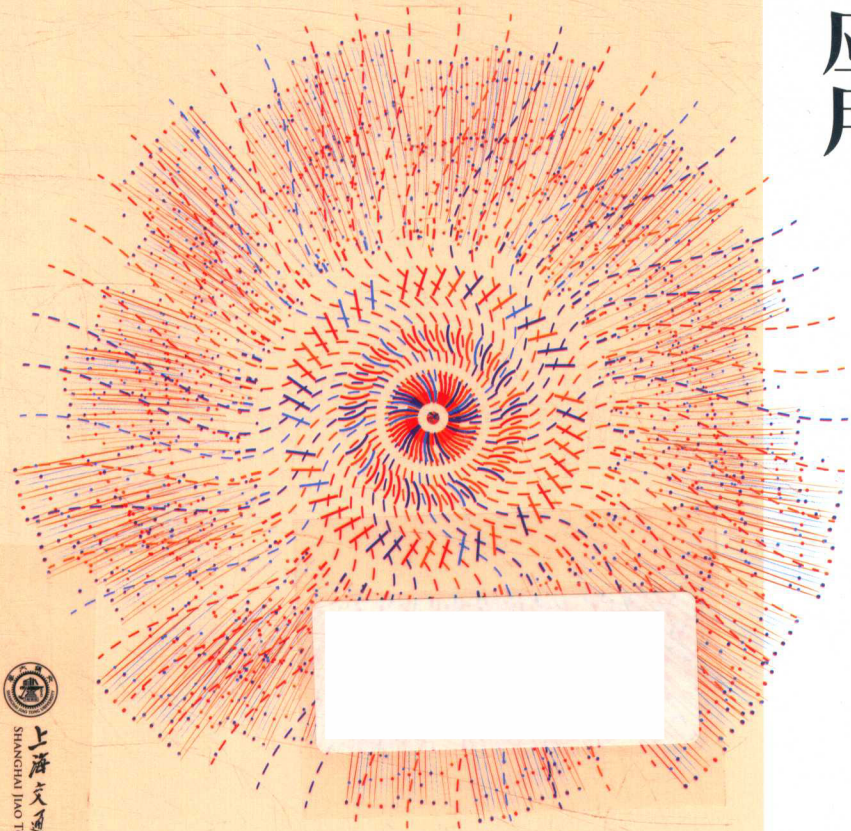
外国语言文学前沿研究丛书

基于语料库的中国学习者英语特征研究及应用

甄凤超 著



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

外国语言文学前沿研究丛书

基于语料库的中国学习者英语特征研究及应用

甄凤超 著

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

图书在版编目(CIP)数据

基于语料库的中国学习者英语特征研究及应用/甄凤超著. —上海:上海交通大学出版社,2019

ISBN 978-7-313-21838-4

Ⅰ.①基… Ⅱ.①甄… Ⅲ.①英语—语言学习—研究—中国

Ⅳ.①H319.3

中国版本图书馆 CIP 数据核字(2019)第 182768 号

基于语料库的中国学习者英语特征研究及应用

著　　者:甄凤超

出版发行:上海交通大学出版社　　地　　址:上海市番禺路 951 号

邮政编码:200030　　电　　话:021-64071208

印　　制:江苏凤凰数码印务有限公司　　经　　销:全国新华书店

开　　本:710mm×1000mm　1/16　　印　　张:14.75

字　　数:247 千字

版　　次:2019 年 10 月第 1 版　　印　　次:2019 年 10 月第 1 次印刷

书　　号:ISBN 978-7-313-21838-4

定　　价:88.00 元

前　言

学习者语料库是国内语料库语言学研究的一个重要内容，从建设初期就达到了国际同类研究的领先水平。1990 年，比利时新鲁汶大学的 S. Granger 教授发表了世界上第一个大型学习者语料库 ICLE，时隔不久，中国也建成了首个大型的学习者英语语料库，即 CLEC，之后又相继建设了一批各具特色的学习者语料库，并触发了大量的相关研究。国内学习者语料库研究的蓬勃发展，一方面得益于我国语料库语言学研究前辈们敏锐的学术洞察力，另一方面也得益于我国拥有庞大的外语学习群体。从建设 CLEC 到今天，学习者语料库研究风雨兼程二十载，经历了从最初的萌芽到后来的发展壮大，也曾遭遇过瓶颈期。目前，学习者语料库研究需要创新，机遇与挑战并存。但作为一种重要的语言研究以及语言教学研究资源，学习者语料库在未来一定会继续发挥其重要作用。

笔者长期从事学习者语料库研究，直接参与了 CLEC 与 COLSEC 的开发建设，并基于这些语料库，开展了一系列理论探索与实证性研究。把这些研究整理汇总成论文集的初衷，一方面是为了重新审视这些研究，根据相关最新文献对研究进行补充与完善，并增加一些新的研究话题，另一方面，把这些原本相对独立的研究按照主题汇集起来，可以更加系统全面地了解我国学习者的英语水平和语言特征，并探讨我国学习者语料库研究的现状、问题以及未来发展方向，为从事相关研究的人员提供有价值的参考信息。

该论文集包括上篇、中篇、下篇三大部分。上篇关于国内学习者语料库研究概况，共分两章。第 1 章综述了国内语料库语言学近四十年的成长历史，分别考察了语料库语言学在外语教学、二语习得、翻译、普通语言、自然语言处理等相关领域的应用。第 2 章则聚焦于国内学习者语料库研究的现状以及未来发展的趋

势。通过梳理文献，归纳出在学习者语料库研究中出现的新特点，如增加跨学科的研究视角、引介新的语言学理论、注重口语语料库研究等，并在此基础上提出一些新的议题，如开发教学语料库、建设学习者专门用途英语语料库、从短语学视角分析学习者语言等。中篇主要基于CLEC和COLSEC，采用数据驱动的方法，从多个维度系统分析中国学习者的英语特征，共有九章。前三章基于CLEC，从语义序列、配价型式、搭配配价的视角集中分析了中国学习者的英语笔语特征，后面六章基于COLSEC，从词汇覆盖率、词汇知识、口语中的预制性语块、话语标记语、反馈语、打断等角度，深入细致地分析了中国学习者的英语口语特征。下篇主要探讨学习者语料库在外语教学中的应用，共有三章，分别介绍了数据驱动学习的理念和方法、学习者语料库在外语教学中的应用途径、如何将配价型式和搭配配价应用于外语教学。后记针对语料库语言学在发展过程中出现的一些不同声音和思潮展开深入探讨，从三个方面阐释了语料库语言学在变化中应有的坚守。

学习者语料库为研究学习者语言和外语教学提供了极其重要的数据资源，本书从不同视角对学习者语料库展开理论探讨和实证分析，但仍难免挂一漏万，尚祈学界贤达指正。

甄凤超

2019年8月

目　录

上篇　国内学习者语料库研究概况

中篇　中国学习者英语口笔语特征多视角分析

下篇　学习者语料库与外语教学

上 篇

国内学习者语料库研究概况

20世纪90年代末，由广东外语外贸大学的桂诗春教授和上海交通大学的杨惠中教授共同主持建设了中国第一个学习者英语语料库(CLEC)，开启了国内学习者语料库研究。之后，国内又连续建设了一批颇具规模的学习者语料库，如中国大学生英语口语语料库(COLSEC)、中国学生英语口笔语语料库(SWECCL)、中学英语教育语料库(MSEE)、国际外语学习者英语口语语料库-中国子语料库(LINSEI-China)等。这批语料库触发了大量的学习者语言研究。二十年风雨兼程，学习者语料库研究正经历着从萌芽起步到不断发展壮大的过程，目前已成为国内语料库语言学研究的一个重要方向，每年都有一定数量的基于学习者语料库的研究报道发表。

然而，我们也应当清醒地看到，目前学习者语料库研究正处于发展中的瓶颈期：已有的语料库数据相对陈旧，需要更新，并且缺乏历时监控语料库；语料库类型较为单一，需要开发新类型的学习者语料库；传统的中介语对比分析需要拓展新的视角和维度，对数据的分析不能只是停留在对语言错误、多用或者少用现象的描述，也要把学习者语言看成是一种自然发生的语言变体，从母语迁移、习得规律、跨语言以及跨文化的角度，深入系统地分析学习者语言；学习者语料库在外语教学中的应用不能只局限于为教学提供诊断式信息，而要拓宽思路，扩大应用范围，并且要充分考虑学习者的学科专业，建设一批具有特定学科文类特征的学习者语料库，服务于特定学科教学的需求；学习者口笔语语料库的发展仍然不均衡，但随着语音识别技术、人工智能技术、自动语音转写技术的发展，学习者口语语料库的建设与研究一定会取得更快的发展。

本书上篇包括两章综述。第1章宏观把控了语料库语言学在中国的成长与发展，以语料库语言学在外语教学、二语习得、翻译、普通语言学、话语分析等领域的发展和现状为主线，评述国内语料库语言学研究的主要特征、成就、问题和发展前景。第2章则聚焦国内学习者语料库研究，通过述评近二十年一些公开发表的文献，描绘国内学习者语料库研究发展的轨迹和特点，找到存在的问题，并在此基础上，探索未来一些有价值的研究课题，以期给相关研究者以启示。

第1章 语料库语言学在中国的成长与发展

1.1 引言

20世纪80年代早期，上海交通大学杨惠中教授组建研究团队，策划建设中国第一个大型电子语料库，即交大科技英语语料库(JDEST)，启动了语料库语言学在中国的发展。风雨兼程四十载，国内语料库语言学研究队伍不断壮大，研究不断深入，研究水平不断提高，不仅建成了一批可与国际同类语料库相比的专门语料库、学习者口笔语语料库、平行语料库、可比语料库，研究话题也从词汇、语法和词典学扩展到语言教学、二语习得、翻译、自然语言处理、话语分析、认知语言学、功能语言学乃至抽象的理论语言学等广泛的领域。越来越多的具有不同学术背景的研究者都积极加入语料库语言学研究的队伍中来，刮起了一股强劲的“语料库风”，而国内外不同学术领域的交流与合作也给中国的语料库语言学研究开辟了更为广阔的发展空间。本章首先将回顾国内语料库建设的三个主要阶段，然后以语料库语言学研究类型为主线，综述语料库语言学研究在中国的发展动态，找到发展中存在的问题，并对未来发展趋势作出展望。

1.2 国内语料库建设的三个主要阶段

我们按照每个阶段建设的语料库的特点，将国内语料库建设大致划分出三个主要阶段：

第一个阶段始于20世纪80年代早期，其主要特征是建设专门用途语料库(specialized corpus)，主要有上海交通大学建设的JDEST语料库(库容为100万

形符)和中国石油大学广州分部创建的广州石油英语语料库(库容为411 000形符)。这些是我国乃至世界范围内的第一代大型电子语料库,是我国在20世纪80年代对早期计算机语料库建设所做出的重要贡献,受到国际学术界的广泛关注。基于这些语料库产生的索引大全与词频表(如《石油英语频率词典》,祝启波著),为全国大学英语教学改革尤其是对大学英语教学大纲的制定、研究专门用途英语词汇以及语言对比和教材编写提供了数据资源。此外,国内同时期还建成了一批汉语语料库,如汉语现代文学作品语料库(527万字,武汉大学)、现代汉语语料库(2 000万字,北京航空航天大学)、中学语文教材语料库(106万字,北京师范大学)和现代汉语词频统计语料库(182万字,北京语言学院)。进入20世纪90年代后,汉语语料库的库容进一步扩大,如北京语言文化大学建立了一个约5亿字的中文语料库,清华大学建立了7—8亿汉字的语料库,等等。这些汉语语料库主要应用于中文信息处理研究,如汉语的切词和消歧。

第二个阶段开始于20世纪90年代中后期,主要以建设学习者英语语料库为主。代表性语料库包括广州外语外贸大学和上海交通大学联合建成的中国学习者英语语料库(CLEC, 100万形符),上海交通大学、河南师范大学和解放军外国语学院联合建成的中国大学生英语口语语料库(COLSEC, 70余万形符),南京大学建成的中国英语学生口笔语语料库(SWECCL, 200万形符)和华南师范大学建成的中学英语教育语料库(MSEE, 450万形符)。此外,还有华中科技大学的硕士写作语料库(MWC, 12万形符),华南师范大学的国际外语学习者英语口语语料库-中国子语料库(LINSEI-China, 10万形符)。这些语料库收集了学习者学习英语时所产生的中介语,因此也被称为中介语语料库,为研究学习者中介语的语言特点和语误现象以及二语习得提供了真实可靠的数据。

第三个阶段开始于21世纪初,是我国语料库建设全面发展的时期。随着建库技术的不断完善,各种类型的语料库如雨后春笋般迅速发展起来。首先,建成或在建一批汉英平行语料库,如中国科学院计算技术研究所的汉英双语语料库(20万句对,提供网上查询服务),北京大学汉语语言学研究中心的CCL汉英双语语料库(233 589句对),北京大学计算语言学研究所的BABEL汉英平行语料库(20万句对),东北大学的英汉双语语料库(100万句对),哈尔滨工业大学的英汉双语语料库(50万句对),北京外国语大学中国外语教育研究中心的通用汉语对应语料库(约3 000万汉字/英文词),南京大学的南大—国关平行语料库,外语教学与研究出版社的英汉文学作品语料库,国家语言文字工作委员会语言文

字应用研究所的计算机专业的双语语料库，上海交通大学的汉英双向平行语料库、莎士比亚多译本平行语料库、汉英会议口译语料库、当代英汉/汉英法律平行语料库等一系列平行语料库以及燕山大学的《红楼梦》中英文平行语料库，等等。这些语料库触发了一系列相关研究，如语言对比研究、翻译研究、双语词典编撰研究、教学研究和机器翻译研究。其次，建成或在建一些特殊英语语料库，如解放军外国语学院的军事英语语料库，河南师范大学的中国英语（China English）语料库等。这些语料库已呈现出良好的应用前景，其中中国英语语料库为研究中国英语变体的形成，研究英语语言从其本族文化的联结中被剥离出来后与其二语或者外语学习者和运用者的本土社会文化及环境的对接提供资源。需要说明的是，上述三个阶段并非楚河汉界、泾渭分明，而是一个彼此交错、相互渗透的过程。

随着语料库的建设与开发，基于语料库的语言研究不断涌现，我们借助中国期刊网的检索平台，统计出与语料库相关的文章数量，如表 1.1 所示。表 1.1 显示了一组 CSSCI 来源期刊发表的有关语料库文章的统计数字：

表 1.1　国内与语料库相关的 CSSCI 期刊文章数量统计表

发表时间（年）	篇数
1982—1985	8
1986—1990	10
1991—1995	35
1996—2000	87
2001—2005	321
2006—2009	759
2010—2018	904

从时间维度上看，语料库语言学研究最近二十年的发展尤为突飞猛进，研究话题也从词汇、语法和词典学扩展到包括语言教学、二语习得、翻译、自然语言处理、话语分析、认知语言学、功能语言学，乃至理论语言学等更为广泛的研究领域。接下来我们以语料库研究的类型为主线，分析语料库语言学在我国的发展动态和存在的问题。

1.3　语料库语言学研究和外语教学研究

外语教学研究是语料库语言学主要应用领域之一，一直备受国外语言学家

和语言教育家的关注(Leech，1997；Renouf，1997；O'Keeffe，McCarthy & Carter，2007)。而语料库语言学在中国发展的初始阶段就与外语教学有着密切的联系。JDEST 语料库，是中国和亚洲地区的第一个学术英语语料库，基于该语料库提取的学术英语领域通用词汇、技术词汇、次技术词汇等信息，为中国大学英语教学大纲的制定提供了可靠的科学量化依据，为推动大学英语教学改革发挥了重要作用。杨惠中教授提出制定教学词表的"定量分析为主，定性分析为辅"的原则：在通过频数、覆盖率和分布率等主要统计特征确定词表之后，经由定性分析做进一步筛选，筛选的依据包括社会学标准、语言教学标准和语言学标准。基于这一原则，上海交通大学从 JDEST 语料库中提取并编写了含 6 000 个词汇的常用科技英语词汇表，为制定"大学英语教学大纲通用词汇表(1—4 级)"提供了重要参考。JDEST 项目的重要意义在于，它开创了我国外语界进行语料库研究的先河，为世界范围内的第一代语料库建设，尤其是专门语料库建设提供了一套经典的原则、方法和技术范式，被 John Sinclair 和 Geoffery Leech 等誉为东方语料库建设的先驱。20 世纪 90 年代中后期国内开始有学者提出把语料库应用到外语教学实践中(谢应光，1996；郭杰克，1997；何安平，2001)。如郭杰克(1997：5)指出："语料库所提供的信息不仅为我们编写教学大纲、教学词表和教材提供了客观和可靠的依据，它还为我们进行外语教学提供了新的思路"。卫乃兴(2007b)在《John Sinclair 的语言学遗产中》一文中也指出，Sinclair 一直对语料库语言学与语言教学的结合秉持积极的立场和态度，并提出了学习者在学习过程中需要掌握的关键技能：①将话语切分为有意义成分的能力；②区分向心式结构与离心式结构的能力；③使用语言对语言认识、讨论、重组的能力；④释义的技能。这些观点的提出均是基于语言描述的研究成果，折射出语料库语言学理论的立场，既具挑战性也有可行性。濮建忠和卫乃兴(2000)在探讨词汇与语法的关系时指出，每个词汇都有其语法，词汇的意义和结构之间存在着极为密切的关系，从而主张从词汇出发，以词的核心用法为中心设计英语教学内容。这与 Sinclair & Renouf(1988)提出的词汇大纲的思路和设想是一致的。他们认为英语教学的重点应放在语言中最常见的词形、其核心用法模式以及典型组合。这种设想的提出，无疑是对传统外语教学中词汇与语法处于相对独立状况的一个突破。2010 年 9 月 24—25 日，首届广外英语语言学论坛在广东外语外贸大学成功举办，其间，举行了题为"语料库语言学与外语教学"的高层论坛，由桂诗春、冯志伟、杨惠中、何安平、卫乃兴、李文中、梁茂成等国内知名语料库语言学专家

参加，就语料库语言学与外语教学等主题展开了互动讨论，专家发言内容刊登在《现代外语》2010 年第 4 期上。如杨惠中先生在讨论中指出："语料库语言学本身就是实践性、应用性很强的一门学科……外语教学证实语料库语言学的重要应用领域，语料库语言学以实际使用中的语言事实作为研究对象是一种着眼于语言语用的研究方式，因此跟语言教学有着直接的关系……语料库研究成果可以应用在教学大纲的设计中，为确定教学内容、制定教学目标提供坚实可靠的决策依据"（参见桂诗春、冯志伟和杨惠中等，2010：422）。

至于如何将语料库应用到实际的外语教学中，Sinclair 提出师生可以直接进入语料库资源，通过观察词语索引和扩展语境，自我发现和归纳语言型式。教师也可以按照教学需求自建语料库，或对现有语料库资源进行深加工研究，应用到语料库辅助的语言教学中。根据我们的不完全统计，自 2000 年以来，CSSCI 来源期刊发表的有关语料库和教学研究的文章多达千余篇，特别是从 2009 年开始，文章发表数量每年都保持在 70 余篇。话题涉及基于语料库的外语教学的理论基础以及各种应用性研究，如基于语料库的词汇、语法、语篇、翻译等外语教学模式。由于这方面的文章数目较大，我们拟以语料库应用于外语教学中的直接程度为线索，对相关文献进行梳理，总结特征，分析潜在问题和困难。

数据驱动学习（data-driven learning，简称 DDL）是一种把语料库数据和检索技术直接应用于外语教学的方法。该方法由英国伯明翰大学的 T. Johns 教授（Johns，1991a）提出，其主要思想是引导学生基于语料库中大量真实语言数据，通过观察、描述、归纳语言使用现象，自我发现语言规律。在国内，李文中和濮建忠（2001）较早介绍了数据驱动学习的方法，探索了语料库索引技术在外语教学中的应用。他们提出了数据驱动学习的三种基本实现手段：其一是开发独立的 DDL 软件，把语料库索引行以及词汇练习一同打包；其二是与其他学习材料结合起来，针对语篇中词汇和搭配制作基于语料库索引的交互式练习，利用网络技术供远程课堂或局域网网络教室使用；其三是利用词语索引进行课堂实时演示，通过教师的参与和指导进行语言学习。甄凤超（2005a）也讨论了语料库数据驱动外语学习的思想、方法和技术。由卫乃兴、李文中和濮建忠等人承担的国家社会科学基金项目"语料库与多媒体技术在外语教学中的应用"（02BYY016）首次在国内实现了语料库资源的在线查询、检索和免费共享，实现了 4 个语料库的 700 多万词的文本资源 KWIC 网络在线转换和实时传送与共享，开发了数据驱动学习系统。以沈阳师范大学白志刚（2009）为核心的团队利用语料库和

Wiki平台等手段，进行了对英语专业高年级自主学习模式的探索。有研究者对比数据驱动学习模式与传统外语教学模式，结果显示前者能有效提高学生词汇水平，并对培养学生自主学习能力和研究性思维有一定促进作用（俞燕明，2009）。显而易见，语料库数据驱动的外语学习和网络多媒体技术为外语教学提供了一条崭新的思路。另外，许多功能强大、操作简单的语料库检索软件的出现使得语料库技术更为方便地进入外语教学课堂。除了一些商业软件如Wordsmith Tools，Concordance之外，许多检索软件可以免费下载使用，如AntConc等。语料库数据驱动的外语学习强调大量自然语言数据的输入、真实语言学习环境的营造、自主学习能力的培养，而这些都是传统外语教学规约式课程和内省式数据所无法比拟的。但是，必须同时指出，DDL教学模式和实践仍处于探索阶段，加上目前国内外语教学条件的限制，语料库数据与技术在外语教学中的应用仍有漫长的路要走。

语料库间接应用于外语教学的一条重要途径，是外语教师对语料库资源，特别是学习者语料库进行深加工研究，用于诊断式外语教学。国内目前建成的具有一定规模的学习者语料库有CLEC，COLSEC，SWECCL和MSEE等。利用英语本族语语料库和学习者语料库，开展中介语对比研究，概括、描述学生中介语的使用特征并诊断其错误，对症下药，保证了教学的效果。

国内一批平行语料库的建设以及基于平行语料库的翻译研究成果，触发了大量的基于语料库的翻译教学研究。例如，王克非（2004）较早地介绍了双语平行语料库在翻译教学上的用途，主要体现在三个方面：①对某一检索词或短语提供丰富多样的双语对译样例；②为常用结构提供多种双语对译样例，供讲解和仿习；③提供丰富的可随机提取的一本多译作为对照参考。近些年，基于语料库的翻译教学研究呈现多学科交叉的趋势，如人工智能以及大数据挖掘技术在语料库翻译教学中的应用（王克非、刘鼎甲，2017；徐琦璐，2017），从认知语言学的视角谈语料库翻译研究及教学应用（如：郭高攀、廖华英，2016；胡开宝、李晓倩，2016）。

1.4 语料库语言学与二语习得研究

我国自20世纪90年代中后期开始建设学习者语料库，近年来发展迅速。到目前为止，已建成如下几个颇具规模的学习者语料库：

（1）CLEC是国家哲学社会科学“九五”规划项目，由广东外语外贸大学桂

诗春教授和上海交通大学杨惠中教授主持建设，建库的目的就是为了对学习者的英语进行深入的研究，语料库光盘版已于 2003 年由上海外语教育出版社出版。CLEC 收集了包括中学(St2)、大学英语 4 级(St3)和 6 级(St4)、英语专业低年级(St5)和高年级(St6)在内的 5 种学习者语料共 100 多万词，并对言语失误进行标注。

(2) COLSEC 是国家哲学社会科学“九五”规划项目，由上海交通大学杨惠中教授主持，上海交通大学、洛阳解放军外国语学院、河南师范大学等高校的教师和研究人员参加。语料采自从 2000 年至 2004 年全国大学英语考试口语考试的实景音像资料，涵盖教师—学生型晤谈、学生—学生型自由讨论、教师—学生型讨论共三类题材的内容，较为全面地反映了中国大学生在英语交谈活动中的语音语调特征、词汇语法结构特征、话语结构特征与会话策略使用情况，全库总容量为 723 299 个形符。该语料库应当是我国国内第一个可与国际同类语料库比较的学习者英语口语语料库，为研究我国大学生的英语口语能力建造了较为坚实的数据资源。关于 COLSEC 语料库建设与研究的专著《中国学习者英语口语语料库建设与研究》(杨惠中、卫乃兴主编)已由上海外语教育出版社出版。

(3) SWECCL 由国家“211 工程”二期子项目中国学生英语口语语料库(简称 SECCL)和教育部人文社科项目“中国大学生英语写作能力发展规律与特点”的数据库中国学生英语笔语语料库(简称 WECCL)两个子项目组成，语料库的设计总规模为 200 万形符，其中 SECCL 口语子库和 WECCL 子库各为 100 万形符，项目由南京大学主持、与外语教学与研究出版社合作共同开发。SECCL 语料收集了从 1996 年至 2002 年英语专业四级的口试录音资料。WECCL 语料主要收集国内 9 所不同层次的高校英语专业的 1—4 年级学生的命题作文，文体为议论文，也有少量的记叙文和说明文。此外，两个语料库皆标注词性码。

(4) MSEE 是 1998 年广东省高等学校电化教育“五个一百工程”的立项课题之一，由华南师范大学何安平教授主持。该语料库包括 180 万形符的英语教材语料(含初中、高中和大学英语课本教材)，120 余万形符的国内外英语课堂(大、中、小学)130 节课的教学实况语料(配录音或录像)和 150 万形符的国内初、高中、大学学生英语口语和书面语语料。该语料库光盘版于 2000 年由广东教材音像出版社出版。

另外，还有一些研究者根据不同研究目的自建的学习者语言语料库，包括英语之外的其他语种的学习者语料库。

这些语料库的建成触发了一系列对中介语和二语习得的研究。其中，CLEC 的应用最为广泛，据中国期刊网的不完全统计，基于 CLEC 的研究论文多达百余篇。研究多数通过对比学习者英语与本族语者英语，进行学习者英语特征分析和错误分析。研究话题涉及搭配、类联接、语义韵、句型、语篇、语体等特征描述，成果斐然，对二语习得研究以及英语教学研究具有重要的参考价值。到目前为止，COLSEC，SWECCL 和 MSEE 的应用价值也日益彰显，相关研究成果也逐渐增多（罗颖，1999；冯友，2005；甄凤超，2005b；卫乃兴，2007a；王立非、文秋芳，2007；郑群，2011；杨江锋，2013；徐璐，2015）。与经典二语习得研究相比，语料库证据支持的中介语研究具有以下特点：

（1）采用自下而上的研究方法。自下而上，即从真实语言使用的数据出发，依赖语言数据的频数或者概率信息，通过提取（extraction）—观察（observation）—概括（generalization）—解释（interpretation）的研究过程，描述语言事实，抽象语言学理论。语料库证据支持的方法，与经典二语习得研究不同，不预设研究假设，不受太多的理论模型约束，所使用的数据也较之经典的二语习得研究惯用的内省数据和诱导数据（主要通过问卷调查、实验研究和个案追踪的方法获得）更加客观、真实和丰富。

（2）采用对比研究的方法。Granger（1998）提出的基于语料库的“中介语对比分析”（contrastive interlanguage analysis）是近年来兴起的二语习得研究的方法。该方法采用语料库研究的基本技术手段和方法，通过对比本族语与中介语、不同母语背景的中介语、相同母语但不同习得阶段的中介语在一系列维度上的相关数据，概括出学习者和本族语者的差异、中介语的模式和学习者行为趋势，发掘中介语的非本族语特征，并探索引起这些特征的背后原因（卫乃兴，2006a）。

需要指出的是，上述提及的多数学习者语料库从建成到现在，已有十余年时间，有的甚至超过了二十年，语料相对陈旧，不足以呈现目前我国学习者的语言特点。我们需要收集新的学习者语料，按照统一的建库标准，建设一批新的学习者语料库，一来可以与上一代语料库进行纵向比较，二来也可以从共时的角度，系统分析目前我国学习者的语言使用特征。

1.5 语料库语言学与翻译研究

基于语料库的翻译研究是近年发展起来的又一重要研究领域。英国学者 M. Baker 教授（1993）最早创设了“语料库翻译学”的研究范式，突破了传统翻译