

YULIAOKU ZAI YINGYU JIAOXUE  
ZHONG DE YINGYONG YANJIU

# 语料库在英语教学中的应用研究

侯 静 著

第 10 卷



重庆大学出版社

YULIAOKU ZAI YINGYU JIAOXUE  
ZHONG DE YINGYONG YANJIU

# 语料库在英语教学中的应用研究

侯 静 著

重庆大学出版社

## 内容提要

《语料库在英语教学中的应用研究》内容理论联系实际、重点突出。全书分五章,分别为绪论、语料库应用于英语教学的基本理论、语料库应用于英语教学的背景和实验方案、语料库在英语课堂教学中的实际应用、语料库在英语教学中应用的反思。

本书适合英语专业本科生、研究生阅读,也可作为语言学及应用语言学方向研究人员的参考用书。

### 图书在版编目(CIP)数据

语料库在英语教学中的应用研究 / 侯静著. -- 重庆:  
重庆大学出版社, 2019.12

ISBN 978-7-5689-2031-5

I. ①语… II. ①侯… III. ①语料库—应用—英语—  
教学研究 IV. ①H319.3

中国版本图书馆 CIP 数据核字(2020)第 017116 号

## 语料库在英语教学中的应用研究

侯 静 著

策划编辑:范 琪 鲁 黎

特约编辑:向 璐

责任编辑:范 琪 版式设计:范 琪

责任校对:王 倩 责任印制:张 策

\*

重庆大学出版社出版发行

出版人:饶帮华

社址:重庆市沙坪坝区大学城西路 21 号

邮编:401331

电话:(023) 88617190 88617185(中小学)

传真:(023) 88617186 88617166

网址:<http://www.cqup.com.cn>

邮箱:fxk@cqup.com.cn(营销中心)

全国新华书店经销

重庆市正前方彩色印刷有限公司印刷

\*

开本:787mm×1092mm 1/16 印张:5.75 字数:106千

2019年12月第1版 2019年12月第1次印刷

ISBN 978-7-5689-2031-5 定价:48.00元

本书如有印刷、装订等质量问题,本社负责调换

版权所有,请勿擅自翻印和用本书

制作各类出版物及配套用书,违者必究

## 前 言

计算机技术的创新化发展和网络技术的不断成熟引发大数据时代的到来,在此基础上,现代数字信息技术也有很大程度的发展,并且在教育教学领域中的应用不断增加。在英语教学中,语料库是新时代教学中不可或缺的一种教学工具,与大数据技术高质量结合的新时代语料库具有检索速率快、促进教学和科研智能化发展、语言具有真实性等优势。同时,在教学方面,新时代语料库能通过具体的量化指标对英语教学进行专业化、精准化的指导,能够有效促进教师英语课堂教学质量的提升;在科研方面,利用英语语料库,研究人员能将制定好的语法现象进行大规模量化,同时能够相对精准地对英语语言理论、语法规则以及理论假设进行检测和验证。

本书从语料库应用于英语教学的“基本理论”“背景和实验方案”“课堂应用和反思”这几个维度对英语教学中语料库的作用进行研究,并进一步提出英语教学中语料库应用的局限性,希望能够为英语教学中语料库的应用提出更多有价值的理论参考。

由于著者水平有限,书中难免存在不足和疏漏之处,敬请广大读者批评指正。

著 者

2019年5月

# 目 录

<b>第一章 绪论</b> .....	1
第一节 语料库简述 .....	1
第二节 语料库在英语教学中的应用研究概况 .....	1
<b>第二章 语料库应用于英语教学的基本理论</b> .....	7
第一节 语料库的发展及主要类型 .....	7
第二节 语料库应用于英语教学的理论基础 .....	12
第三节 语料库应用于英语教学的特点 .....	19
<b>第三章 语料库应用于英语教学的背景和实验方案</b> .....	21
第一节 语料库应用于英语教学的实施背景 .....	21
第二节 语料库应用于英语教学的实验方案 .....	22
第三节 语料库应用于英语教学的预期教学效果 .....	26
<b>第四章 语料库在英语课堂教学中的实际应用</b> .....	29
第一节 语料库与词汇教学 .....	29
第二节 语料库与语法教学 .....	65
第三节 语料库与写作教学 .....	70
第四节 语料库与翻译教学 .....	74
第五节 语料库与专门用途英语教学 .....	75

<b>第五章 语料库在英语教学中应用的反思</b> .....	77
第一节 语料库应用于英语教学的意义 .....	77
第二节 语料库应用于英语教学存在的问题 .....	80
第三节 语料库在英语教学中的前景和方向 .....	81
<b>参考文献</b> .....	84

### ★第一节 语料库简述

语料是人们运用语言时自然发生的书面或口头的语言材料(语言素材)的集合,即现实生活中的语言,而语料库顾名思义就是存放语言的“仓库”,但这个“仓库”里的语言不是杂乱无章地堆放在一起的。语料库是一个由大量在真实情况下使用的语言信息集成的、专供研究和使用的资料库,是按照一定的目的与方式建立起来的、大规模储存这些语言材料的“仓库”。对语料库的研究与应用其历史之长远可以追溯到中世纪。那时,在西方,以字母顺序排列伴随索引语境的索引词表就用于对宗教文件进行文本分析。而语料库语言学作为一种方法论则出现在 20 世纪 60 年代,它是一门独具特色的语言研究学科,以语料库为手段来研究语言。现代语料库语言学(modern corpus linguistics)则是 20 世纪中后期兴起的一门语言研究科学。它是指专门对大规模储存于计算机里的语料库进行研究的学问。这种电脑语料库是一个由大量在真实情况下使用的语言信息构成的、可供计算机检索的、专门作研究使用的巨型语料库。它以容量大、语料真实、检索快捷准确等独特优势,在现代语言学研究和语言教育中发挥着越来越大的作用。

### ★第二节 语料库在英语教学中的应用研究概况

#### 一、国外研究综述

语料库语言学一般被认为是一门新兴的语言研究科学。其发展历经四个阶段:第一阶

段,19世纪末,人工检索和手工卡片制作阶段。当时的语法书和词典通过这些卡片及手工检索来编写。第二阶段,20世纪60—80年代,跨国语料库开始建设。第三阶段,20世纪90年代末期,早期的语料库内容被更新;另外,开放式、滚动式的语料库建立。第四阶段,从21世纪初,具体是2005年起,电子网络语料库建立。运用互联网等现代信息技术,建立了海量、动态、共享的语料库。

具有代表性的语料库有:20世纪80年代末,英国朗曼出版公司建立的朗曼学习者语料库(LLC),属于书面语语料库,容量为1 000万词,是最早出现的较大规模学习者语料库。剑桥学习者语料库(GLC),由英国剑桥大学出版社组织建立,词汇容量超过2亿条。1990年比利时 Sylviane Granger 教授带头创立了国际英语学习者语料库,简称ICLE,容量为200万词。ICLE的语料组成广泛而多元,国内外研究者、学习者参与人数多,被看作是最具影响力的语料库之一。其他的书面语语料库还有匈牙利英语学习者语料库(JPU)、瑞典英语学习者语料库(USE)、日本英语学习者语料库(JEFLL)、波兰英语学习者语料库(PELCRA)及美国英语学习者语料库(MELD)等。

语料库可以研究的内容很多,其中包括语料库的英语教学。Elena Semino 教授、Vaclav Brezina 和 Dana Gablasova 博士分别作了“Corpus-based Stylistics”“Corpus Linguistics and Second Language Acquisition”“Statistics in Corpus Linguistics”等报告,在这些报告中,他们阐释了如何利用语料库开展各种应用性研究,探讨了如何制作各种类型的语料库、语料库统计及利用语料库数据进行语体分析等内容。

J.Leech 在1998年提出“教学是科研的自然延伸”。J.Sinclair 是把语料库实际应用到教学中的先行者,他在2003年的著作 *Reading Concordances* 中阐述了如何使用语料库开展语言教学创新实践活动。

Tognini-Bonelli 指出:语料库语言学并不是一个真正意义上的科学研究领域,只不过是语言研究提供了一种方法论。同时语料库又给语言学的研究提供了新的哲学思路,是介于理论和方法论之间的一种东西。

## 二、国内研究综述

我国语料库的建设开始于20世纪80年代,起初只是用语料库对汉语词汇进行统计并研究。20世纪90年代之后,各种类型的语料库逐步被建立,语料库方法被大量、经常地使用,有关语料库的种种问题都逐渐出现并研究。20世纪90年代末到21世纪初是语料库深入开发

和应用的时期。

### (一) 建设各类语料库, 提倡语料库资源的共建和共享

著名语言学家、英语教育家、广东外语外贸大学桂诗春教授被认为是应用语言学主要创始人之一, 为我国的语料库应用语言学发展作出重大贡献。桂诗春教授形容当时的窘况说: “最初只能开设一门笼统的课, 浅浅涉及应用语言学、心理语言学和社会语言学等方面的内容。相关的研究资料严重匮乏, 基本是摸着石头过河。”中国英语学习者语料库(CLEC)由桂教授和来自上海交通大学的杨惠中教授一起主持, 国内十几所院校共同参与建设。英语专业的学生语料、大学英语四六级考试语料以及中学生学习英语的各种书面语资料都被收集在这个语料库中, 词汇超过了100万条, 很难得的是还对中国学生的言语失误进行了标注。桂诗春教授迫切地提出了以下问题, 呼吁研究者们应该集中起来一起解决这些问题: 语料库语言学该在什么层次上开设? 它的基本内容是什么? 如果在大学开设有关语料库的课程, 它应由哪些核心课程组成? 桂教授还提出语料库语言学是依赖资源的。如果想要推进语料库语言学在我国快速发展, 应当合作, 进行资源共享。

北京外国语大学中国外语教育研究中心专职研究员梁茂成教授于2014年作了“语料库及其在外语教学与研究中的应用”学术报告, 明确指出语料库可以用于理论或应用研究, 从而摆脱了之前学术界对其独立学科地位的争论。梁茂成教授在报告中讲述了语料库的重要性。他认为在语言教学中, 语料库能够避免老师在课堂上的单向主导地位, 大量真实自然的语言可靠度更高。在语言研究领域中, 语料库数据的充足性、真实性、高效性及可复制性, 增加了研究的科学阐释力。梁教授还指出语料库与英语教学密切相关, 在教学大纲的制定、教材的编写、英语学习、英语学习测验等方面都可以运用语料库。他还介绍了国际语料库语言学界的最新研究, 其中包括搭配计算方法、差异检验方法、新媒体分析方法、语料库深度加工等多方面。梁教授指出, 语料库应用于外语教学除了遵循教育教学规律为首要原则外, 还应该加强语料库资源的共建与共享。

### (二) 提出语料库可用于理论或应用研究

华南师范大学教授、博士生导师何安平教授提出了“教学加工”的理念, 运用语料库来进行外语教学, 带来了一系列从课堂模式到教材设计的改革。首先, 大型语料库为英语教育提供新鲜的教学资源和教学手段。在外文学院领导和有关部门的大力支持下, 何老师与人合作研制了《中学英语教育语料库》光盘, 将常规的印象图文与大型语料库结合为一体, 实现了多

界面转换。其次,运用语料库手段于课堂教学,实现了大信息量、多维演示的立体教学和人机互动的学习模式,体现了“以学生为主体的、注重实践的、注重创新能力的培养”的现代教育思想。运用语料库进行教学,教师既亲身参与对话料的探索,又要引导学生如何使用语料库进行语言调查分析和归类,从而为学生提供了一个发挥主动性、发挥创造能力的空间。因为学生一上机就面对着庞大的语料库而置身于研究者的位置,他们要亲自动手通过使用电脑软件在语料库中检索、观察、思考,把看似纷繁杂乱的语料现象转变为有规律、有意义的信息符号。

国内外研究语料库的众多专家学者都在思索何教授的“教学加工”理念,思索语料库语言学是如何应用于英语教学的:其中包括不同层次的英语教学大纲如何根据语料库来设计,面向不同专业、不同层次的教材如何根据语料库来编写,课堂教学的活动如何根据语料库来设计,教学的实施及教学活动后的评价如何根据语料库进行等。何教授指出语料库应用于外语教学的发展趋势:一要大力建设面向各个行业、各种层次的多种用途的外语学习者语料库;二要发展语料库,进一步完善语料库,例如,可以在建设中尝试把音频与视频相连接,甚至建设音频与视频可以同时共现的语料库;三要考虑把关于语料库的检索、使用、应用及研究纳入师范教育学生的必修课程,或将其纳入在岗教师的继续教育课程。

教育部语言文字应用研究所研究员、计算机语言学家冯志伟,专门从事语言学和计算机科学的跨学科研究。冯志伟教授认为语料库最大的贡献就是改进了英语教学,尤其是英汉双语语料库更具有优势:研究者或英语教师通过双语语料库检索到教材的原资料,这能帮助学生从整体上感知真实语言资料,培养学生的语感,也能让教师脚踏实地编写出实用的外语教材。然而之前的英语教材是由教师根据自己的语感或前人经验来编写的,既费时又费力,还往往满意度不高。可若是根据英汉双语语料库编写教材,就可以在巨大的英语资源库里比较容易地搜集到可用的素材,工作效率大大提高了;另外教材的真实性高,也更具实用性和针对性。

### (三)从教师培训做起

研究语料库的重要方面之一就是在进行外语教学中使用语料库,这是上海交通大学杨惠中教授对研究语料库的看法。语料库研究的对象是在实际中真实出现过的语言,语料库研究法是一种研究语言实际运用的研究方法,因此语料库与英语有着密切的关系。实际语言交流的各个方面包括词语、句型、文体、语义等都在语料库中进行了深刻的探讨和标注。语料库可

以在外语教学大纲的设计中使用,语料库丰富的内容为教学内容的确立及教学目标的制定都能提供真实可靠的依据。语料库教学是指由教师根据学生的水平和层次开发建设一个有针对性的、有目的性的微型语料库;运用有关的检索软件,学生可以快捷地搜索到词汇或短语等;通过对比分析外语与母语,学生可以自己尝试探索英语词语的意义、用法及规律等。

用这种语料库学习方法,学生自己先直接地接触到英语自然、真实的使用情况,再在教师的引导下,通过对语境的观察,自己动脑分析、与母语对比,与之前学习掌握了词汇知识对比、自己归纳或者与同伴讨论,发现语言的使用规则,进行探索式学习。杨教授还指出:过去几十年间,许多英语专业的学生通过研究语料库获得了硕士或博士学位,但是在实际应用语料库服务于教学方面做得还不够,语料库还没能完全发挥应有的作用。

所以大力发展语料库教学,就要先从培养教师开始。要对教师进行语料库知识的培训,培训教师学会使用语料库、研究语料库。在应用语料库教学前,先让教师在语料库中学习语言,这对教师的专业和个人发展都有极大的意义。

### 三、国内与国外研究综述评析

#### (一)蓬勃发展,多方面参与探究

关于语料库的研究在国际上被分为三个方向:一是设计开发各种用途的语料库,也包括开发建设有关语料库的软件。这些往往依靠语言学及计算机软件制作的专业人士协作完成。二是在语料库中进行学生学习外语的习得研究,其中外语的语音、语法、词汇等都囊括在内。三是针对语料库的应用研究,包括利用语料库资源编纂外语词典,编写各种类型、各个方面的教材,随着科技的进步,还出现了电子产品。

根据国内外的研究可知,“语料库语言学是语言学科中飙升得最快的学科之一,语料库语言学是一门年轻的学科,语料库应用是一种很有意义的活动,近年来,这种活动魅力四射,吸引了很多人参与”。在教学方法、学习方式及有关词汇、语法、写作、翻译等有关教学的各个方面,语料库都有积极的应用。因为运用语料库可以引导学生进行探索式和发现式学习,所以许多研究者运用语料库进行案例教学,他们指导学生对检索出的真实语言进行独立的分析和归纳、找出规律,然后从中得出英语语言的相关特征和规律,从而指导英语学习。

#### (二)教师由“消费者”转变为“合作研发者”

以前,国外语料库研究主要是宣传语料库,对语料库在教学中应用的可能性和意义进行广泛介绍;而如今的语料库研究进一步发展到如何有效地与其他学科的理论与方法相融合的

阶段。在这一阶段,研究语料库教学中的理论问题需要思考解决、相关软件工具需要不断地更新或升级、课堂活动的设计和教学方法都需要再实践、再修正等。在国内,虽然语料库成为改进教学的新资源和新技术被众多专家学者所认可,专家们都认为用语料库研究外语教学很有价值,但是针对语料库的英语教学还没能真正地进入普通的课堂教学,只是处于实验阶段。2010年,“教学与语料库国际研讨会”在捷克召开。这次召开的已是第9届会议,会议提出应该转变这些观念:由专家学者宣讲语料库的用途转变为由代课老师提出语料库可以用来做什么,然后由专家和教师一起研究怎样解决语料库在教学上真实存在的问题。外语教师要由语料库及“语料库产品”的使用者或“消费者”转变为“合作研发者”。

### (三)与课堂教学实际相结合成为趋势

现阶段我国英语语料库的建设和应用仅仅局限于国内大学生尤其是英语专业大学生,缺乏对中小学、中职及高职学生的英语语料的搜集,还不能全面反映我国英语学习者的整体语言面貌;语料库的建设和应用方面缺少与国外及地区之间的合作,欠缺与语料库相关的计算机基础知识和相关软件的开发。要想真正与国际研究接轨,还需要研究语料库的专家和一线外语教师及研究语料库的计算机人才合作。高职英语教学有其特点,不能完全照搬本科英语专业的语料库。语料库语言学的发展要求把语料库和课堂教学活动结合在一起,这项研究成为语料库应用于英语教学的趋势。本研究要注重培养学生自主型与探究式的学习,充分利用语料库的特点,将大量真实生动的语言以特别明显的方式呈现给高职学生。学生通过阅读教师挑选出的难易适中的英语资料,经过自主学习、与他人交流讨论等方式获得语言的意义、用法、搭配等。这种方法,可以加深学生的记忆,实现有效学习。本研究针对高职学生,用语料库辅助教学,在课堂中运用软件给学生演示如何学习是很有必要的。

## ★第一节 语料库的发展及主要类型

### 一、语料库的发展阶段

自 20 世纪 50 年代开始,计算机技术突飞猛进,互联网开发和应用也日新月异,为语料收集、存储、加工、检索和提取提供了重要的资源和技术支持。语料库的发展大致经历了三个阶段:

#### (一) 第一阶段——18 世纪初到 20 世纪 50 年代

第一阶段为语料库发展的初始阶段,是计算机语料库出现以前的各种手工语料库的发展时期。尽管依靠手工所收集的语料量小,样本代表性不强,进行语料处理和分析工作浩繁,但早期的语料库仍然在以下领域起到了重要的作用:圣经与文学研究、词典编撰、方言研究、语言教育研究(包括儿童语言习得研究,拼写、语言教学以及比较语言学研究等)。

这一阶段代表性的语料库有:

#### 1. 为词典编撰、语法研究而收集的语料库

牛津英语词典(*Oxford English Dictionary*)1928 年,引证 400 万条,卡片 1 100 万张。

韦伯斯特新国际词典(*Webster's New International Dictionary*)1961 年第三版,新旧引证 1 000 多万条。

#### 2. 为教学目的而编制的书面语料库和词表

陈鹤琴《语体文应用字汇》,商务印书馆,1928 年。进行过两次统计,第一次统计使用了

六种材料,包含 554 478 个汉字的语料,得出不同汉字 4 261 个;第二次统计使用包含 34 818 个汉字的语料,得出与 4 261 个汉字相异的汉字 458 个。

### 3. 为语言调查而收集的方言库

19、20 世纪英、美等国都做过大型的方言调查,调查的结果形成几个大规模的方言库。在我国,运用语料的研究方法可以追溯至周秦,如我国汉语方言的第一部著作《輶轩使者绝代语释别国方言》(简称《方言》)是扬雄经过 27 年的艰苦努力而写成的。

### 4. 布朗语料库(Brown Corpus)

布朗语料库(Brown Corpus)是 20 世纪 60 年代中期构建的一个具代表性的平衡语料库。尽管由现在的理论及技术水准看来,布朗语料库的资料及平衡方式略显粗糙,可是这个语料库一直是(英语)平衡语料库的标准,甚至到了 20 世纪 80 年代新构建的英语平衡语料库如 LOB(Lancaster-Oslo/Bergen, 英国英文)及 London-Lund(英语口语),都还遵循布朗语料库的架构,足见这种平衡语料库在各种语言学研究上有其不可取代的价值。

## (二) 第二阶段——20 世纪 60 年代到 80 年代

因为计算机技术的飞速发展,第二阶段的语料库建设以电子语料库的兴起为主要特征,语料库的容量不断扩大,种类不断增多。

这一阶段代表性的语料库有:

### 1. 英国国家语料库(British National Corpus)

英国国家语料库是由英国牛津出版社、朗文出版公司、钱伯斯-哈洛普出版公司、牛津大学计算机服务中心、兰开斯特大学英语计算机中心以及大英图书馆等联合开发建立的大型语料库,于 1994 年完成。英国国家语料库以来源广泛的书面语言和口语为样本,收录了 1 亿词的电子资源,用以呈现 20 世纪后期以来的英式英语,涉及口语和书面英语。它采用国际通用标准化标注体系 SGML,使用三级赋码标注,使标注错误率由 3% 减少到 1%。在应用方面,该语料库既可用其配套的 SARA 检索软件,也可支持多种通用检测软件,并可直接进行在线检索。

### 2. 朗文语料库(Longman Corpus Network)

朗文语料库是商用语料库,建于 20 世纪 80 年代,由三个大语料库组成,这三个语料库分别是朗文/兰开斯特英语语料库(Longman/Lancaster English Language Corpus, LLELC)、朗文口语语料库(Longman Spoken Corpus, LSC)、朗文英语学习者语料库(Longman Corpus of

Learners' English, LCLE)。朗文语料库建立的主要目标之一是编纂英语学习词典,为外国人学习英语服务,规模达到 5 000 万词级。

### (三) 第三阶段——20 世纪 90 年代至今

第三阶段是语料库建设、研究与应用的迅猛发展阶段。因为功能更为强大的计算机系统软件的开发与应用,第二代超大型计算机语料库开始出现,库容扩大到至少要以千万甚至亿计。

这一阶段代表性的语料库有:

#### 1. COBUILD 语料库

COBUILD 语料库的建设初衷是为编纂新的英语词典提供语料支持。该语料库是由英国伯明翰大学和 Collins 出版社合作建成的,因此被称为 COBUILD(Collins Birmingham University International Language Database)或 Birmingham 语料库。COBUILD 语料库的结构和内容是根据一定的原则设计的,以满足学习者、教师和其他使用者的需要,同时体现其对于研究当代英语的价值。

COBUILD 语料库的创新之处,主要表现在两个方面:一是规模上的突破;二是将语料库建设与具有商业目的的基于语料库的词典编纂、语法书编写和语言教学课程开发相结合,语料库语言学研究从纯学术走向了实用。

#### 2. Longman 语料库

Longman 语料库主要由 Longman/Lancaster 英语语料库(LLELC)、Longman 口语语料库(LSC)和 Longman 英语学习语料库(LCLE)三个语料库组成。

#### 3. BNC 语料库

英国国家语料库(British National Corpus,简称 BNC)是英国牛津出版公司、朗文出版社、钱伯斯-哈洛普出版公司、牛津大学计算机服务中心、兰卡斯特大学英语计算机中心以及大英图书馆等联合开发建设的大型语料库,于 1994 年完成。BNC 是一个平衡语料库,书面语言和口语样本来源广泛,规模为 1 亿词次,用以呈现 20 世纪后期以来的英式英语。

#### 4. ICE 语料库

国际英语语料库(International Corpus of English,简称 ICE)是伦敦大学学院 SEU 项目负责人 Sidney Greenbaum 于 1998 年提议建设的。该语料库由 20 个子语料库构成,每个子语料库的规模为 100 万词,语料采集自 18 岁以上、接受过以英语为教学媒介语言的正规教育、受

教育的程度至少为初中水平的成年人使用的材料。口语和书面语言的抽样时间段为 1990—1993 年。建立国际英语语料库的目的,是对世界范围内英语的不同地域变体进行比较研究,探索英语的地域变体在词汇、句法等层面的差异。

与 BROWN、LOB、SEU 等语料库相比,ICE 主要有以下几个特点:①口语语料所占的比例高于笔语语料。BROWN 和 LOB 两个语料库中的语料全部来自书面语言,SEU 语料库中 50% 的语料来自书面语言,而 IEC 中只有 40%。②在 IEC 语料库的 300 个口语文本中,普通人之间的谈话数量多于 SEU 语料库,后者收集的口语语料多为各种广播。③ICE 中不再有在谈话者不知情的情况下录音的口语文本,反映了研究方法的伦理变化。④ICE 中独白的子类更多,而 SEU 语料库中演说和评论两个子类组成了大部分的独白文本。⑤为了保持转写的速度和一致性,口语文本没有韵律标注。⑥ICE 书面语言语料的各部分虽与之前的语料库有类似之处,但是语料库代表性的实现方式不同,书面语言文本中虚构性语料所占的比例不同就是一个例子。

## 二、语料库的类型

语料库的类型可以从以下几个层面来划分:第一类,从语言种类来划分,可分为单语语料库、双语语料库以及多语语料库;第二类,从语料信道来分,可分为书面语料库和口语语料库;第三类,从载体媒介来分,可分为印刷文本、电子文本、口语转写等;第四类,从应用层面来分,可分为通用语料库和专用语料库。如我国语言学家桂诗春和杨惠中建立的中国英语学习者语料库就属于专门语料库;第五类,从语料处理方面来分,可分为生语料库和标注语料库。经过标注和赋码的语料库使得语料库数据的分析更加系统和精确,也便于对特殊数据信息的提取和处理。

因为语料库是为不同的研究目的建立的,其必然有多种类型,以代表各种语言、语言变体或文类。常见的语料库类型主要有:

### 1. 通用语料库 (general corpus)

通用语料库广泛采集某种语言的口语和书面语材料,取样时充分考虑该语言的主要社会变体、地域变体、行业变体及使用场合之间的平衡;文本取自一个或多个国家,尽可能最好地代表一种语言的全貌。通用语料库的规模一般较大,常常达到上亿词次或数亿词次,既为语言学习或翻译提供参考资料,也可用作与专用语料库进行比较的依据,因此也被称为参照语料库 (reference corpus)。著名的英语通用语料库主要有英国国家语料库、英语文库以及更早

的 LOB 和 BROWN 等一代语料库。

### 2. 专用语料库 (specialized corpus)

与通用语料库相反,专用语料库只收集某种特定领域的文本,如报刊中的社论、地理教科书、某个特定学科的学术文章、讲座、日常谈话、学生作文等。专用语料库对文本的专业程度没有任何限制,但限制文本的选择范围。例如,文本的时间框架、谈话的交际情景、文本所涉及的话题等。专用语料库通常代表某个特定类型文本的特征,用来调查研究特定领域内语言的特点或编写专门领域的工具书。比较著名的专用语料库有体现非正式英式英语的剑桥与诺丁汉英语谈话语料库 (Cambridge and Nottingham Corpus of Discourse in English, CANCODE) 和密执安学术英语口语语料库 (Michigan Corpus of Academic Spoken English)。

### 3. 平行/双语语料库 (parallel/bilingual corpus) 和多语语料库 (multilingual corpus)

平行/双语语料库中的语料来自两种不同的语言,一种语言是另外一种语言的译文。因为实际应用的需要,平行语料库中的两种语言必须对齐,即一种语言的某个语言单位和另一种语言相应的语言单位要对齐。多语语料库中的语言样本取自多种语言,并经过对齐处理。平行语料库是翻译者实践成果的宝库,对翻译研究和机器翻译具有重大意义。

### 4. 学习者语料库 (learner corpus)

学习者语料库中的语料取自某一语言的非本族语学习者,可分为口语语料库和笔语语料库。建立学习者语料库的目的是研究学习者语言之间以及学习者和本族语者语言之间有哪些方面不同,通常采用和本族语者语料库相比较的方法。国际上颇有影响的学习者语料库是 20 世纪 90 年代建立的英语学习者国际语料库 (International Corpus of Learner English, ICLE), 库中收集了不同国家英语学习者的作文。可用于与之相比较的英语本族语者作文语料库是鲁汶英语本族语者作文语料库 (Louvain Corpus of Native English Essays)。

### 5. 教学语料库 (pedagogic corpus)

教学语料库是由学习者接触到的语言样本组成的语料库,通常由教师或研究者视教学需要建立。库中的语料可以包括学习者使用的所有教科书、简易读物和音频材料等。例如,教学语料库可用于收集学习者在不同的语境下接触到的词汇或短语的例证,来强化他们的学习意识,也可用于与自然发生的语言样本组成的语料库进行比较,来确定学习者接触的语言是自然的、有用的。

### 6. 共时语料库 (synchronic corpus)

共时语料库是相对历时语料库而言的,代表某一时期的语言或文类。例如, Brown 语料库