

中国学习者语料库研究的多元化转向 ——《基于语料库的中国学习者英语特征及应用》述评

张正坤, 阿力亚·艾尼

(上海理工大学 外语学院, 上海 200093)

摘要:《基于语料库的中国学习者英语特征及应用》由上海交通大学甄凤超教授著述, 囊括了其本人从事学术研究以来开展的一系列有关语料库研究的理论探索和学习者英语特征的实证性研究。该书从语义序列、配价型式、搭配配价、词汇覆盖率、词汇知识、预制语块、话语标记语、反馈语、打断等角度, 对我国学习者语言特征开展了深入细致的研究。这是一本实证研究和教学应用无缝对接的佳作, 研究结论对外语教学具有重要的参考价值和指导意义。述评旨在对该书主要内容及其概括性、示范性、多样性、全面性、引领性、启发性等特色予以评介。

关键词: 中国学习者英语特征; 学习者语料库; 外语教学; 评介

中图分类号: H 319

文献标志码: A

文章编号: 1009-895X(2023)03-0258-06

DOI: 10.13256/j.cnki.jusst.sse.220507237

The Diversification Turn of Chinese Learners' Corpus Study —A Review of *A Corpus-based Study on the Features and Applications of Chinese English Learners*

ZHANG Zhengkun, ALIYA Aini

(College of Foreign Languages, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: *A Corpus-based Study on the Features and Applications of Chinese English Learners* is written by Professor Zhen Fengchao of Shanghai Jiaotong University. The book covers a series of the author's theoretical exploration of corpus studies as well as empirical studies on the English features of Chinese learners in his academic research. The book discusses Chinese learners' English features in terms of semantic sequence, valency pattern, collovalency, lexical coverage, vocabulary knowledge, prefabricated chunk, discourse markers, reaction tokens, and interruption. The work seamlessly integrates empirical research and pedagogical application, and its findings are of great reference value and guidance for foreign language teaching. This paper intends to review the main content and highlights the comprehensive summary, valuable guidance, multifarious perspectives as well as its enlightenment to foreign language teaching.

收稿日期: 2022-05-07

基金项目: 2022年度教育部人文社科青年项目(22YJC740002); 2021年外语教材研究院青年项目(2021SH0012)阶段性成果

作者简介: 张正坤, 男, 硕士研究生。研究方向: 语料库翻译、话语分析。E-mail: 1753722960@qq.com

通信作者: 阿力亚·艾尼, 女, 副教授。研究方向: 语料库翻译、话语分析。E-mail: aliya809@hotmail.com

Keywords: Chinese learners' English features; learners' corpus; foreign language teaching; review

自20世纪80年代初始,上海交通大学杨惠中教授及其团队策划建设了JDEST语料库,语料库语言学在我国已发展壮大40余载。学习者语料库研究(Learner Corpus Research, LCR)是语料库语言学研究的一个重要内容,Granger于1998年建成第一个英语学习者语料库(International Corpus of Learner English, ICLE)^[1],自此学习者语料库研究开始在国内方兴未艾。《基于语料库的中国学习者英语特征研究及应用》(以下简称《应用》)是一本文集,于2019年10月出版,是上海交通大学出版社《外国语言文学前沿研究丛书》中的一部。该书共收录12篇学术论文,包含综述部分及作者数年来基于学习者语料库开展的一系列理论探索和实证性研究。《应用》聚焦于国内学习者语料库的研究现状、问题及趋势,基于CLEC和COLESC两大中国学习者语料库,从多个维度和视角,全面系统地分析我国英语学习者的英语笔语和口语水平,以此总结出学习者的普遍特征,为我国的英语教学提供有益参考。

一、内容概述

《应用》分为上、中、下三篇和附录部分。上篇包含两章,第1章立足于国内语料库语言学的成长与发展,第2章预测国内学习者语料库的研究热点及发展方向。中篇是主体部分,涵盖9个章节,均为作者本人多年来基于学习者语料库进行的实证性研究。其中前3章分析了中国英语学习者的笔语特征,后6章分析了中国英语学习者的口语特征。下篇共有3章,聚焦学习者语料库在外语教学中的具体应用以及对学习者语料库前景的展望。

(一) 国内学习者语料库研究现状综述

第1章简要回顾了国内语料库如何从初始走向繁荣,并将其发展历程分为3个阶段。接着考察了语料库语言学在不同领域的研究,以此为主线,综述了国内语料库语言学的发展,并对未来的研究热点进行展望。作者预测语料库语言学未来的研究热点将聚集在翻译、话语分析、二语习得、语言教学以及自然语言处理等诸多领域。此外,语料库语言学驱动的短语学研究是未来的主流研究方向,且历

时语料库语言学研究 and 基于语料库的批评话语分析研究也将日益蓬勃。

第2章聚焦国内语料库语言学研究现状及发展趋势。通过述评近20年公开发表的文献,作者概述了国内学习者语料库研究发展的历程和特点,指出国际上学习者语料库研究理论、视角和方法都有所更新,国内研究亟待深化,并建议未来的研究重点应随着时代进步而侧重于跨学科的研究视角、引介新的语言学理论及加强口语研究。最后阐述了未来学习者语料库研究的新议题应着重于教学语料库的开发和应用、学习者专门用途英语语料库研究以及从短语学视角分析学习者语料库。

(二) 中国学习者英语笔语特征多视角分析

第3至9章是基于CLEC和COLSEC语料库展开的一系列实证性研究,在数据驱动的基本框架下,采用理论与数据、描述与解释、检索与对比有机结合的研究方法,从多方位多角度分析归纳了中国学习者的英语特点,旨在为我国英语教学提供有益参考。

第3章采用语义序列(semantic sequences)的视角,以语法词of为例,分析了CLEC语料库中语法词的短语学特征。本研究选取CLEC中大学英语四六级作文及英语专业高年级作文作为研究语料,并根据学生的英文水平,将其分成S1、S2、S3三个子库。作者通过检索发现学习者最常用的of短语模式,并且概括总结了学习者在写作中常用n of n短语模式所表达的含义,进而对比了不同英文水平学习者使用n of n短语模式所表达含义的异同。研究结果证明了语法词可以作为短语学特征研究的切入点 and 基于小型语料库、利用频数信息研究短语学特征的可行性,对未来相关研究的开展具有借鉴意义。

第4章结合型式语法(pattern grammar)和配价语法(valency grammar),引介了一套新的配价型式描写体系,并以动词consider为例,描写了该动词在CLEC中的配价型式特征。型式语法反映了语料库短语学词汇和语法不可分的一元论^[2],而Teubert认为它仅关注了词与词性这一语法范畴,忽略了型式内部成分之间潜在的逻辑语义关系^[3]。为了应对这一问题,作者提出一套新的配价型式描

写体系,将型式语法和配价语法有机结合,在保留词与词性这一语法范畴的基础上又增加配价语法的句法功能。研究者使用 Wordsmith Tools,在 CLEC 中对动词 consider 检索,并以 Reichardt 基于 BNC 制作的 consider 配价结构表^[4]为参照,进而归纳总结出动词 consider 的所有配价结构。研究发现,中国英语学习者在写作中倾向于使用 consider 的二价结构和三价结构,且学习者使用动词 consider 时误用现象频发。此外,作者通过对比分析发现,中国学习者和英语母语者使用动词 consider 配价结构的表达含义以及 consider 与其后面两个名词短语的语义逻辑方面呈现了较大的差异性。这套新的配价分析描写体系阐述了词汇和语法一体化的语言特性,代表着数据驱动短语学研究的新尝试。

第5章在前一章配价型式的基础上,提出了搭配配价(collovalency)的概念及其统计方法,并以动词 consider 为例在 CLEC 中调查分析了该动词的搭配配价特征。搭配配价是传统配价语法的简化和升级,它舍去了传统配价语法对价的数量和性质的严格界定,仅延用了传统配价结构描述体系中的部分语法范畴,强调词与词之间的搭配力^[5]。对 CLEC 和 BNC 语料库中学习者最常用的动词 consider 的配价结构进行检索和对比,发现中国英语学习者与英语本族语者关于动词 consider 的配价结构的使用型式基本一致。此外,作者考察了词与其配价结构之间的搭配力,从中得出中国学习者口语交际中最具典型性的动词 consider 配价结构。与传统基于频数信息的搭配研究相比,搭配配价力是一种更为科学、精确的检验方法;通过搭配力的计算,可以得出配价结构的典型性,即配价力越高,典型性也越高。该研究不仅展现了配价分析在语料库语言学研究中的应用前景,也提及了配价结构和搭配配价在英语词汇教学中应用的具体途径和方法,这对外语词汇教学有所裨益。

第3至5章基于 CLEC 语料库,从语义序列、型式语法和搭配配价三个视角考察了中国学习者的笔语特征。研究结论不仅为外语教学提供参考,也为未来开展同类研究带来启迪,如以语法词为切入点开展语料库驱动的短语学研究、基于学习者语料库开展搭配配价研究等。

(三) 中国学习者英语口语特征多视角分析

从第6章起,本书焦点开始转向学习者口语特征研究。

第6章从词汇覆盖率(lexical coverage)的角度,采用词目统计和对比分析等方法,比较了 COLSEC 和两个本族语者口语语料库的词汇覆盖率和词汇使用情况,以探究中国学习者在口语交际中的词汇使用特征。本研究使用 Wordsmith 软件对 COLSEC 和 CLEC 语料库及 BNC 和 ICE 语料库的口语部分进行了词目统计。数据显示, COLSEC 中词目总量明显低于 CLEC 和 BNC,反映出中国学习者在口语交际中词汇量输出有限,这在一定程度上限制了口语的交际效果。此外,作者分别总结了4个语料库中最常用的100个词目,并对其进行了相关性分析。数据表明,中国学习者和本族语者在口语交际中的实词使用有很大差异,本族语者更善于用种类繁多的话语标记词(discourse marker),且搭配形式层出不穷,相比之下,中国学习者实词使用稍显拙略,这导致中国学习者的口语会话缺乏交际性。故,未来的英语教学需加强针对英语会话常用词汇的教学。

第7章基于 COLSEC,采用“主题词与关键主题词”的研究方法,对中国英语学习者口语交际能力中的词汇知识进行了考察,研究结果对于外语词汇教学具有启发意义,且可为 CET-SET 及其他类型的口语考试提供反馈信息。基于心理语言学中图示理论(schema)和心理词库(mental lexicon)的概念,作者认为学习者在口语交际中会使用话题图示触发该话题的心理词汇,使之最终成为具体词汇。首先,本研究将语料按照话题分类,并就某一话题随机选择3个不同口语水平的学习者的语料。此外,选用 ICE(口语文本)作为参照语料库,运用 Wordsmith 在观察语料库 COLSEC 中筛选出主题词和关键主题词,即观察语料库中出现频率比在参照语料库中显著偏高的词,并借助 Wordsmith Tools 的词汇索引和搭配功能,分析了主题词和关键主题词的语义特征,并以网络图示的方式描绘了主题词和不同词语场的关系,这便于直观分析学习者心理图示和心理词汇触发过程。研究揭示了话题图示和心理词库与学习者口语表达能力呈显著正相关,且不同英语水平的学生在话题图示和词汇知识上存在显著差异。这一章的价值在于指导教师在外语词汇教学环节中应着重培养学习者建构话题图示的意识以及运用丰富的词汇知识准确表达意义的能力。

第8章介绍了预制语块(prefabricated chunk)使用与外语口语能力关系的实证性研究。基于 COLSEC 语料库,这部分内容从学习者口语交际的

流利度、准确度和恰当性三个角度, 考察了语块使用与外语口语能力的关系。目前, 学界对于语块识别尚无一套完整的方法, 且口语语料中的停顿、重复、非言语声音等加大了语块提取的难度。作者采用机器识别和人工识别相结合的方法, 利用 Word-smith 的自动提取词丛功能初步识别语块, 在此基础上进行人工复筛并将结果与英语本族语参照语料库进行对比, 以确保筛选的语块为本族语者所接受。该研究综合考察了流利度、准确度和恰当性与语块的关系, 三者的量化需要参考多个参数, 其中流利度的量化最为复杂, 通过量化语速 (SR)、话步均长 (MLR)、停顿长度 (PL)、流利度分数 (FS)、非言语声音 (NVS) 和修改重复 (FRR), 在此基础上借助 SPSS 统计软件对语块与 3 个量化指标分别进行了相关性分析, 充分考察了语块使用与外语口语能力的关系。该章研究的意义在于验证了语块使用与外语口语能力的正相关性, 启示英语口语教学应加强语块意识的培养和语块教学。

第 9 章基于 COLSEC 语料库, 以 well 为例, 探究了中国学习者英语口语中话语标记词的使用特征。早期关于话语标记词 well 的研究主要基于两种研究范式——理论驱动和语料库数据驱动型, 前者关注其心理凸显性 (psychological salience), 后者关注其出现频数。本研究采取语料库数据驱动的研究范式, 探讨了不同口语水平及中国学习者和本族语者使用话语标记词 well 语用功能的异同, 并发现了中国学习者习得话语标记词 well 语用功能的规律性特征。启示教师在今后的口语教学中, 应在遵循学生习得规律的基础上, 加强口语标记词的输入和训练, 着重提高学生的语用能力。

第 10 章基于 COLSEC 语料库, 从使用反馈语 (reaction tokens) 的类型和反馈语出现的位置两个维度出发, 对中国学习者英语会话中反馈语的使用特征进行了实证性研究。值得一提的是该章还对男性和女性学习者常用反馈语类型和频数进行了比较研究。结果表明, 女性学习者比男性使用更多的反馈语, 但差异并不显著, 作者认为是由于为了追求较高的成绩, 男性和女性在考试环境下都表现得更加积极。研究同时发现, 中国学习者反馈语储备较为匮乏, 且反馈语的使用在一定程度上受到汉语习惯的影响, 其主要原因是英语语用环境的缺乏。因此, 在未来的口语教学中, 教师应注重培养学生的语用交互能力, 加强对常用反馈语的训练。

第 11 章基于 COLSEC 语料库, 对比了 3 组不

同口语水平学生交际会话中打断 (interruption) 的异同。学界关于打断的研究多采取定性的研究方法, 或是对个案进行话语分析, 少有基于口语语料库的相关研究。该章通过考察考试环境下听者打断说话者讲话的时间、原因和方式, 对打断的语用功能进行了区分, 并统计了不同类型打断在话轮中的出现频数。本章亮点在于对不同英语口语水平学习者的打断使用进行了对比, 发现口语水平较高的学习者往往能够恰当使用各类打断, 流畅自然地掌握会话的走向。研究证明了打断与学习者的口语交际能力的相关性, 恰当运用打断可以提高会话的交互性和动态性。这一章的意义还在于为 CET 口语测试提供反馈信息, 提供口语测试信度和效度。

第 6 至 11 章基于 COLSEC 语料库, 分别从词汇覆盖率、词汇知识、预制语块、话语标记语、反馈语和打断等角度, 详细调查了我国学习者的口语特征。研究的内容涵盖语言的多个层面, 包括词汇、短语、语法、功能及语用等。总体而言, 其研究结论对外语口语教学及口语测试具有重要参考价值。

(四) 学习者语料库与外语教学

最后三章探讨学习者语料库在外语教学中的具体应用。

第 12 章介绍 Tim Jones 提出的数据驱动学习 (Data-driven learning, DDL) 的思想、主要特征以及具体开展数据驱动学习的方法。DDL 方法具有自主学习、真实语料、自我探索和自我发现、自下而上的学习方式等特点^[6], 挑战了以教师和教材为中心的传统教学理念。在实际教学环节中, 教师可采取 KWIC 索引、词语搭配、类连接、扩展语境、词及“数据驱动”学习课件等手段切实贯彻 DDL 理念。由此, DDL 不仅可以培养学生自主学习的意识, 也利于地道外语的习得。

第 13 章从直接应用、间接应用两个角度对学习者语料库和外语教学的有机结合提出了建议, 并阐释了学习者口语语料库对外语教学的重要意义。直接应用指师生基于大量的语料库数据, 采取语料分析手段, 通过观察、分析和概括语言使用特征来解决教学中的实际难题; 间接应用指将基于学习者语料库的中介语发现应用到外语教学中, 从而教师可以针对学习者的语言特点和学习困难进行有效的课堂干预。最后, 作者指出学习者口语语料库在数量和规模等方面均落后于笔语语料库, 但其在外语教学中的价值不容小觑。

第14章与第4、5章遥相呼应,论述了配价结构及搭配配价在词汇教学中的具体途径和方法。该章首先阐释了语料库语言学的教学价值及当前短语教学,尤其是搭配教学所面临的困境。接着,作者认为配价结构及搭配配价凸显了词汇、结构及意义的共选关系,对英语词汇教学具有借鉴意义。最后,详细介绍了配价结构及搭配配价在教学中的应用,例如教材和工具书的编纂以及课堂教学设计等。

以上3章深入系统地探讨了数据驱动学习方法、学习者语料库、配价结构和搭配配价在外语教学中的具体应用途径及方法,为基于语料库的外语教学提供了新思路。

二、简要评介

总体而言,《应用》一书有如下特点。

(一) 概括性和示范性

《应用》各章都引用了大量的文献,其中篇部分的每一章都是按照先概述前人研究后进行实证分析的框架展开,层层递进,逻辑分明。一般的参考书或指南往往浅尝辄止,仅通过文献堆砌对相关领域研究进行一般性的概述,而该书贵在对各领域当前研究进行批判和反思,这无疑有助于促进语料库语言学学科的健康发展。该书的批判性概述围绕该领域最具代表性的研究、主要的研究范式、常用的研究方法以及语料的选用展开。首先,该书给予语料库语言学视域下的短语学研究高度重视,通过反思前人研究,作者发现当前短语学研究的缺省之处并进行改进和提升,例如第3章通过对语法词of的个案分析,证实了以语法词为切入点开展短语学研究的可行性。其次,《应用》中篇对于研究方法的论述和个案分析,对后人开展学习者语料库的相关研究具有很强的示范性。以学习者会话中的反馈语为例,作者详细论述了如何对反馈语进行分类、如何选择语料以及如何开展定量分析和定性分析。因此,本书也不失为语料库语言学领域的一本上佳的操作手册。

(二) 多样性和全面性

一方面,《应用》采取的研究视角全面。当前,语料库语言学框架下的短语学研究主要考察搭配、搭配框架、形式语法、语义序列、扩展意义单位、

词项、语义韵、词束、意义移变单位、同现词列、配价语法、局部语法等^[7]。《应用》第3至11章分别从以下8个理论视角开展语料库研究:语法词、动词配价结构、搭配配价、词汇覆盖率、词汇知识、语块、话语标记语、反馈语及打断。另一方面,《应用》采取了多维度的对比分析方法:1)本族语者语料与非本族语者语料进行对比,如第4和第6章中,作者选取了BNC、ICE等英语本族语者语料库作为参照语料库,对比发现中国学习者在语言使用时常出现的误用现象,以及在写作和口语表达时的语义偏好和交际能力的差异等;2)同一非本族语者语料库中不同水平和不同性别之间进行比较,如第3、8、9和11章中作者分别对比分析了不同英语水平的学生英语语法词、语块、话语标记词以及打断使用的区别,在第10章中作者对比了男性和女性在对话情境下反馈语使用的异同;3)口语语料和笔语语料之间进行对比,如第6章基于CLEC和COLSEC对比了中国学习者口语和笔语中最常用词目的差异。语料库的对比分析法有助于人们更全面地了解中国学习者的英语特征、语用失误和不足之处。

(三) 引领性和启发性

《应用》宏观把控学习者语料库发展的轨迹和特点,审视了当前语料库语言学研究面临的机遇和挑战,并在此基础上,提出了未来一些有价值的研究课题,极具引领性和导向性。以学习者语料库的开发为例。语料库在外语教学中的作用不容忽视,尽管大多数语料库教学研究强调了英语本族语者语料库在教学中的作用,但Granger认为单一的本族语者语料库不能体现对于学习者有价值的语言知识,故不便于指导教师有目的地开展教学活动^[8]。作者认为教学语料库专注于外语教学而非语言研究,因此倡议大力发展教学语料库和学习者专门用途语料库,这拓展了国内学习者语料库研究的视野,对后续的学习者特征研究和外语教学都具有重要意义。此外,本书关注研究结果对外语教学的启示。《应用》的实证研究中,每一篇都强调了该研究对于外语教学的意义,并对外语教学的诸多环节提出了指导性建议。卫乃兴等认为,学习者语料库的特点在于基于大量真实语料,规避了内省数据和诱导性数据的不足,更利于归纳学习者的普遍语言特征,提供更为真实客观的反馈,而多重语料库证据则能更全面、更深入地反映二语特征^[9]。基于CLEC和

COLSEC 等学习者语料库,该书使用了考试环境下中国学习者的真实语料作为数据源,研究结论具有较高的演绎性,可直接应用于口语教学、词汇教学、教材开发和语言测试等领域,为改进二语教学提供启示。

诚然,本书也存有些许抱憾之处。其一,语料类型较为单一。本书基于 CLEC 和 COLSEC 语料库,采用数据驱动的方法,对学习者英语特征进行分析。然而,CLEC 和 COLSEC 语料库仅包含学习者考试环境下的语料,不能完全反映学习者日常情景下的英语使用特征。例如,该书第 10 章对比分析发现男性和女性学习者在会话中反馈语的使用频数上并未呈现出显著性差异,这在一定程度上归因于考试情景提升了学生参与会话的积极性,故性别因素导致反馈语的使用差异缩小。其二,对学习者笔语特征的研究略显薄弱。该书共有 3 篇实证性研究考察了学习者的笔语特征,而对口语特征的考察有 6 篇之多,不免使读者有顾此失彼之感。此外,作者从语义序列、配价型式和搭配配价的视角集中分析了中国学习者的英语笔语特征,可见重点还是在词汇研究,学习者翻译和写作研究没有涉及。其三,研究结论有待从语用层面深化至认知层面。该书研究对于不同性别、不同母语、不同英文水平学习者在语法、语义、语用等层面存在的共性规律和个性差异进行了考察,意犹未尽之处在于没有追溯其差异产生的原因。如若加入对学习者的认知思维过程的考察,这将对英语课堂教学模式和教学策略的选择大有裨益。

三、结语

学习者语料库为研究学习者语言 and 外语教学提供了极其重要的数据资源,《应用》一书高屋建瓴地描绘了当前中国学习者语料库研究的现状、问题

及未来发展趋势,并从多个角度对学习者的语料库开展理论探讨和实证分析,有利于全面了解我国学习者的英语水平及语言特征。总的来说,该书实证研究和教学应用并举,视野宽阔,内容对于外语教学具有很强的参考价值,是一本不可多得的学习者语料库研究的佳作。

参考文献:

- [1] GRANGER S. *Learner English on Computer*[M]. London: Addison Wesley Longman, 1998.
- [2] HUSTON S, FRANCIS G. *Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English*[M]. Amsterdam: John Benjamins Publishing, 2000.
- [3] TEUBERT W. Sinclair, pattern grammar and the question of *hatred*[J]. *International Journal of Corpus Linguistics*, 2007, 12 (2): 223 – 248.
- [4] REICHARDT R. Valency sentence patterns and meaning interpretation: case study of the verb “consider” [D]. Birmingham: University of Birmingham, 2014.
- [5] 甄凤超, 徐芳红. 搭配配价研究——以学习者英语中的动词 *consider* 为例[J]. *外语与翻译*, 2017, 24 (4): 44 – 50.
- [6] 甄凤超. 语料库数据驱动的外语学习: 思想、方法和技术[J]. *外语界*, 2005 (4): 19 – 27, 40.
- [7] 甄凤超. 基于语料库的中国学习者英语特征研究及应用[M]. 上海: 上海交通大学出版社, 2019.
- [8] GRANGER S. A bird's eye view of learner corpus research[C]// Granger S, Hung J, Petch-Tyson S. *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins Publishing, 2002: 3 – 33.
- [9] 卫乃兴, 陆军. 基于语料库的二语学习研究述评: 范式变化与挑战[J]. *外语教学*, 2018, 39 (5): 47 – 53.

(编辑: 朱渭波)