

国内中介语语料库发展述评

王 宝

河南理工大学, 中国·河南 焦作 454100

【摘要】随着第二语言习得理论的研究深入,有关中介语理论的研究日益丰富,当前中介语语料库的建设无疑是其中的热点话题。在学者提出“汉语中介语语料库”这个概念后,学界便对此开展了不少研究。本文大致梳理介绍了汉语中介语语料库的发展,指出汉语中介语语料库建设过程中显现的特点、当前建设存在的问题以及针对建设的建议,以此介绍当前国内的中介语语料库的发展趋向。

【关键词】中介语; 汉语中介语语料库; 综述

引言

中介语(Interlanguage)指的就是介于母语和学习者要学习的目的语之间的一种特殊的语言系统(Selinker, 1972),是二语习得研究中的重点与热点。汉英中介语产生于母语为英语、目的语为汉语的学习者的学习过程中,是中介语的一种变体——也是中介语的一种具体形式,研究汉英中介语有利于揭示英语为母语的学习者进行汉语作为二语习得的规律,语言教学工作也得以科学地进行。而中介语语料库则是中介语理论研究实践的一种方法,通过建立起中介语语料库的系统研究,方便我们探知二语习得的发展过程。为梳理该领域研究现状,探寻其发展方向,笔者以“中介语语料库”为关键词在CNKI中国期刊全文数据库检索论文后,多方筛选得出一些有参考价值的篇目,本文拟利用检索结果,从中介语理论简介、汉语语料库发展、语料库建设存在的问题、针对语料库建设的建议以及语料库以后的研究发展方向来综述当前国内中介语语料库的研究情况。

1 中介语理论简介

中介语(Interlanguage)的定义最早是由Selinker在1969年提出的,这是第二语言习得理论研究中的一个重要的组成部分。早在1972年Selinker发表了“Interlanguage”,论文中初次引用了中介语的概念,并作了十分详尽的阐述与介绍。与此同时他点出,二语学习者在学习第二语言时,无意识地建立起一个不仅依赖于母语,但又异于母语也异于目的语的一种介于两者之间的语言系统。根据他的定义标准,中介语既可以看成学习者在学习二语时在某一阶段中出现的认知目的语的方式和结果的特征系统,即一种特定、具体的中介语言,也可以视为反映所有学习者在整个第二语言习得过程中认知发生和发展的特征性系统,即一种普遍、抽象的中介语语言体系(Interlanguage Continuum)。他还指出“中介语这个语言系统本身就是一个从阶段到过程的双重系统和庞大体系”,即“母语→中介语→目的语”系统中的一个必然过程。在这个系统中学习者以母语为起点经过中介语达到目的语水平,还指出中介语是必须经历的过程,中介语是第二语言认知中的必由之路。

2 汉语中介语语料库

自1995年国内第一个“汉语中介语语料库”建成后,中国大陆和港澳台地区便相继建成大小不一、规模不等、涵盖内容广泛的各种语料库:“HSK动态作文语料库”、“外国留学生语法信息偏误库”、“汉字偏误标注的汉语连续性中介语语料库”、“东南亚华裔留学生作文语料库”、“美国学生汉语作文纵向语料库”、“外国留学生汉语口语纵向语料库”、“小型外国学生汉语口语语料库”、“外国留学生口语语料库”、“汉语学习者汉字偏误数据资料库”、“韩国留学生汉语中介语语料库”、“多层偏误标注的国别化汉语中介语动态语料库”等。而张宝林,崔希亮曾在2013年提议

建设的“全球汉语中介语语料库”,目前已经建设成1.0版面向全球免费开放使用。

3 语料库建设的特点

张宝林,崔希亮(2013)指出:第一,口语语料库的建设逐渐成为学界热点。第二,多模态语料库的建设也已亟需建设。第三,除针对学习者国别化、多母语背景化的语料库之外,还有针对单一国别、某单一文化背景的学习者的中介语语料库。第四,在对外汉语教学领域的中介语语料库建设的影响与推动下,国内少数民族的汉语中介语语料库建设也已经开始。第五,汉语中介语语料库的建设是在国外语料库语言学的发展和国内外母语语料库建设的影响下开始和发展起来的,但其主阵地始终在国内。第六,学界对语料标注模式的研究逐渐深入。

4 当前语料库建设存在的问题

张宝林(2010)在建设初期,就指出:“语料库数量少且规模小,语料库的语料不全;语料库建设缺乏统一规则与标准,建库实践过于随意;语料库功能不完善;语料标注效率低,标注质量差;各高校之间独立建设的语料库资源不能共享”。

张宝林,崔希亮(2013)在观察一些新建立的语料库后提出评价,指出:汉语的中介语语料库整体间设计水平高低不一;各语料库之间发展不均衡;关于语料库本体的研究贫乏;语言本体研究不足以支撑;大部分语料库建成后不对外开放使用,未能发挥其该有的作用。郑通涛,曾小燕(2016)则认为:汉语中介语语料库建设未能体现跨学科的特点,应多层次吸收利用其他学科,缺乏经典且贴近生活的口语语料资源,语料数据来源存在局限性,缺少建立学习者的历史语料库。

张宝林(2020)之后总结了当前中介语语料库发展的现实情况,指出:汉语中介语语料库的书面语料库和口语语料库之间建设地不够平衡,也就是口语语料库少书面的多;现在还未建成基于多模态理论的语料库;建库时选录与标注语料的方法不够客观不够科学;汉语中介语研究中存在僵化现象。当前语料库的建设与研究亟待新的理论、思路与方法来指导与突破目前建设遇到的瓶颈。对这些问题,学界十分关注,也进行了深入的探讨。

5 针对语料库建设的建议

5.1 库别建设的建议

王韞佳,李吉梅(2001)提出设想建立“汉语中介语语音语料库”的基本框架,试图建立一个专门收录语音层面的中介语语料库。

甘瑞瑗(2005)则提出建立“国别化汉语中介语语料库”,“国别化”就是要“针对不同的国家而实行不同、差别的汉语的教学与研究”。经国别化的汉语中介语语料,为国别化教学大纲的设计、国别化教材的编写、国别化词典的编撰提供了十分可靠的依据,也为汉语作为第二语言的习得研究提供了稳重的多维度基础

数据。

张宝林, 崔希亮(2013)提出建设和研究“全球汉语中介语语料库”,为解决库之间发展不均衡以及兼顾便捷性原则,而设想建设全球统一的语料库,如今已成为现实。

5.2 各库建设的普适性建议与原则

为了建设汉语中介语语料库系统,储诚志,陈小荷(1993)提出建设的目标:语料的内容系统且全面;库规模大,语料数量充足;语料来源散布均匀,代表性强;库信息完备,加工细致;用户界面友好,检索便捷;软件系统具有的开放性和可维护性,使得系统中的语料处理模型和计算机软件可以随时处理、分析新的语料,不断地扩充库容量。

张瑞朋(2012)提出语料的收集要有“真实性”与“连续性”。张瑞朋(2013)再次强调语料库的语料属性必须具有真实性、平衡性。平衡性指国别与层级的平衡性。

肖奚强、周文华(2014)提出汉语中介语语料标注的全面性应该从标注的广度、深度、角度和准确度四个维度来思考,以此来确定语料的类别方便查阅。

周文华(2015)认为汉语中介语语料库要具有多样性,多样性是指语料库内部属性的多样性以及外部属性的多样性。其中,语料库内部属性的多样性主要涉及文本产出者和文本属性两方面。同时,也提到汉语中介语语料库外部属性的多样性,主要是涉及语料库文本类型的多样性和文本时间跨度的多样性两方面。最后,说明了汉语中介语语料库的层次性也可从两个角度进行论述,一是语料库文本的层次性,二是语料库功能的层次性。

李桂梅(2017)提出“影响汉语中介语语料库平衡性的关键因素有:语料采集、语料类型、比例组合”。从“语料采集方面”、“语料属性标注”“确定入库语料”方面采取措施来实现整体语料库的平衡。

颜明,肖奚强(2017)提出建库必须考虑语料的全面性与代表性及语料的同质性与可比性之间的关系;在建设中介语语料库的同时,也应该考虑建设可资对比的学习者母语语料库、汉语本族语者语料库、教学输入语言即教材语言语料库等,对比的视角要求对比的材料具有同质性。

张宝林,崔希亮(2015)提到建库的原则与全面性、可行性、区别性;提出建库的标准流程、收集与录入标准,语料标注的原则。

郑通涛,曾小燕(2016)提出新一代的语料库设计原则:利用词网技术;描述学习者特征;社会语言学特征。新一代汉语中介语语料库的建设原则:真实性原则;广泛性原则;历史性原则;跨学科原则;课内外原则;全媒体原则;无障碍共享原则。

这些原则总结了前人所做的研究与建议,为后面的中介语语料库建设与发展指引着方向。

6 语料库的研究发展方向

张宝林(2020)在第77期《华文教学与研究》中发文:他从有“汉语中介语语料库系统”这个概念,到酝酿与着手阶段算起,汉语中介语语料库的建设前前后后经历快三十年的风雨。期间汉语中介语语料库从无到有,从小到大,从少到多,积极地促进了基于语料库开展的汉语教学的习得研究,“使研究范式从小规模、经验型、思辨性研究演变为基于大规模真实语料的、定性分析与定量分析相结合的实证性研究,使研究结论具备了客观性、稳定性和普遍性的特点”。从“研究需求”、“口语语料选择与转写”、“国别化口语语料库建设的意义”、以及“学习者语料库如何服务于教学应用”等方面,探讨了二

语学习者语料库建设的一些前沿性、迫切性的问题。

7 结语

中介语语料库在我国学界内的应用,将成为一种趋势,当前学术界对于建设各种汉语中介语语料库的呼声此起彼伏,基于语料库的研究也不断增多,无论是在中介语理论发展方面,还是外国留学生在学汉语的语用、认知心理方面,汉语语料库都发挥着不可替代的作用。并且随着趋势,语料库将建设成多种多样,无论是进行国别化建设,还是进行语音、词汇、语法、语用等分类别的建设,还是进行初级、中级、高级这种分层次的设计,都无疑会极大地便捷学界研究中介语,极大地有利于中介语在国内的本土化研究。

同时也希望中介语语料库能按照学界给出的建议与原则来科学合理地建设,避免费时费力以求效率。

参考文献:

- [1] 储诚志, 陈小荷. 建立“汉语中介语语料库系统”的基本设想[J]. 世界汉语教学, 1993 (03): 199-205.
 - [2] 胡晓清. 国别化汉语中介语动态语料库建设理念、实践与前瞻[J]. 山东师范大学学报(人文社会科学版), 2018, 63 (05): 134-156.
 - [3] 李桂梅. “全球汉语中介语语料库”的平衡性考虑[J]. 华文教学与研究, 2017 (02): 46-51.
 - [4] 王轶佳, 李吉梅. 建立汉语中介语语音语料库的基本设想[J]. 世界汉语教学, 2001 (01): 87-92.
 - [5] 肖奚强, 周文华. 汉语中介语语料库标注的全面性及类别问题[J]. 世界汉语教学, 2014, 28 (03): 368-377.
 - [6] 颜明, 肖奚强. 论汉语中介语语料库建设的基本问题[J]. 语言文字应用, 2017 (01): 136-144.
 - [7] 张宝林. 汉语中介语语料库建设的现状与对策[J]. 语言文字应用, 2010 (03): 129-138.
 - [8] 张宝林. 关于通用型汉语中介语语料库标注模式的再认识[J]. 世界汉语教学, 2013, 27 (01): 128-140.
 - [9] 张宝林, 崔希亮. “全球汉语中介语语料库建设和研究”的设计理念[J]. 语言教学与研究, 2013 (05): 27-34.
 - [10] 张宝林, 崔希亮. 谈汉语中介语语料库的建设标准[J]. 语言文字应用, 2015 (02): 125-134.
 - [11] 张宝林. 中介语语料库建设研究新探[J]. 华文教学与研究, 2020 (01): 38.
 - [12] 张瑞朋. 留学生汉语中介语语料库建设若干问题探讨——以中山大学汉字偏误中介语语料库为例[J]. 语言文字应用, 2012 (02): 131-136.
 - [13] 张瑞朋. 三个汉语中介语语料库若干问题的比较研究[J]. 语言文字应用, 2013 (03): 133-140.
 - [14] 郑通涛, 曾小燕. 大数据时代的汉语中介语语料库建设[J]. 厦门大学学报(哲学社会科学版), 2016 (02): 53-63.
 - [15] 周文华. 汉语中介语语料库建设的多样性和层次性[J]. 汉语学习, 2015 (06): 97-105.
 - [16] 于海军. 二语习得中的中介语理论[J]. 西南民族大学学报(人文社科版), 2007 (S1): 103-105.
 - [17] 甘瑞瑗. 国别化“对外汉语教学用词表”制定的研究: 以韩国为例[D]. 北京语言大学, 2005.
- 作者简介:**
王宝(1997-), 男, 汉族, 山东省潍坊市人, 学生, 汉族国际教育硕士, 研究方向: 汉语国际教育。