

国内汉语中介语语料库建设研究现状及未来发展趋势

——以2012~2022年研究为域

朱晓迪

鲁东大学文学院, 山东 烟台

收稿日期: 2024年5月15日; 录用日期: 2024年6月20日; 发布日期: 2024年6月30日

摘要

21世纪以来, 语料库的建设不断繁荣, 汉语中介语语料库研究成为热点话题, 国内汉语中介语语料库建设得到长足发展。本文从CNKI (中国知网)中检索出2012~2022年有关汉语中介语语料库建设的研究成果, 从文献发表情况、研究主题、研究方法三个方面对其进行分析, 探究了国内汉语中介语语料库建设近十年的研究现状并提出未来发展趋势的研究建议。

关键词

汉语中介语语料库, 建设现状, 发展趋势

Current Situation and Future Development Trend of Chinese Interlanguage Corpus Construction

—Taking Study Period from 2012 to 2022 as the Field

Xiaodi Zhu

College of Chinese, Ludong University, Yantai Shandong

Received: May 15th, 2024; accepted: Jun. 20th, 2024; published: Jun. 30th, 2024

Abstract

Since the 21st century, the construction of Chinese interlanguage corpus has been flourishing, and

the research of Chinese interlanguage corpus has become a hot topic, and the construction of Chinese interlanguage corpus has made great progress. This paper retrieved the research results on Chinese interlanguage corpus construction from CNKI from 2012 to 2022, analyzed them from three aspects: literature publication, research topics and research methods, explored the current research status of Chinese interlanguage corpus construction in the past decade, and put forward research suggestions on future development trends.

Keywords

Chinese Interlanguage Corpus, Construction Current Situation, Development Trend

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

语料库是指经过科学取样和加工整理的、具有一定格式和标记的大规模电子文本库。汉语语料库的建设与研发主要趋于两个方向，一是汉语本体语料库，包括了现代汉语，古代汉语与网络社交语言等。二是中介语语料库，也称学习者语料库，是语料库语言学研究中的一个重要分支。随着汉语学习需求日益增长，汉语中介语语料库建设与相关研究已逐渐成为研究热点。

汉语中介语语料库是通过收集以母语为非汉语学习者的自然语料，利用计算机软件工具进行加工整理而成的语言数据库。有关汉语中介语语料库的建设是从 20 世纪 90 年代中期开始的，进入 21 世纪以后得到长足发展。与此同时，也不断推动着汉语中介语语料库应用研究。

2. 研究现状

我国第一个汉语中介语语料库建成于 1995 年，对于汉语中介语语料库的探索已有将近三十年的研究历史。本文以“汉语中介语语料库”和“建设”为主题，在 CNKI (中国知网)数据库中共检索出相关文献 82 篇，并着重语料库建设主题，以 2012~2022 年为封闭域，剔除会议通知、新闻报道等，筛选出 54 篇文献，其中学术期刊 40 篇(核心期刊 19 篇)，硕士论文 8 篇，会议 5 篇(国际 3 篇、国内 2 篇)，中文图书 1 本。

本文拟在前人研究的基础上，从文献发表情况、研究主题、研究方法三个方面对筛选出来的 54 篇文献进行分析，探究国内汉语中介语语料库建设近十年的研究现状并提出未来发展趋势的研究建议。

2.1. 文献发表情况

2.1.1. 发文章统计

2012~2022 年间，有关汉语中介语语料库建设文献每年都有一定数量的发表，可见汉语中介语语料库建设一直是业界研究的话题。

其中 2013 年、2015 年和 2022 年发文章量较多(见图 1)。经查阅资料可以发现，2012 年 12 月北京语言大学对外汉语研究中心主办第九届对外汉语国际学术研讨会；第三届汉语中介语语料库建设与应用国际学术讨论会于 2014 年 8 月 1 日至 3 日在福建福州召开；第六届汉语中介语语料库建设与应用国际学术讨论会于 2021 年 7 月在烟台举办。可以见学术交流活动有利于推动汉语中介语语料库建设研究深入发展。

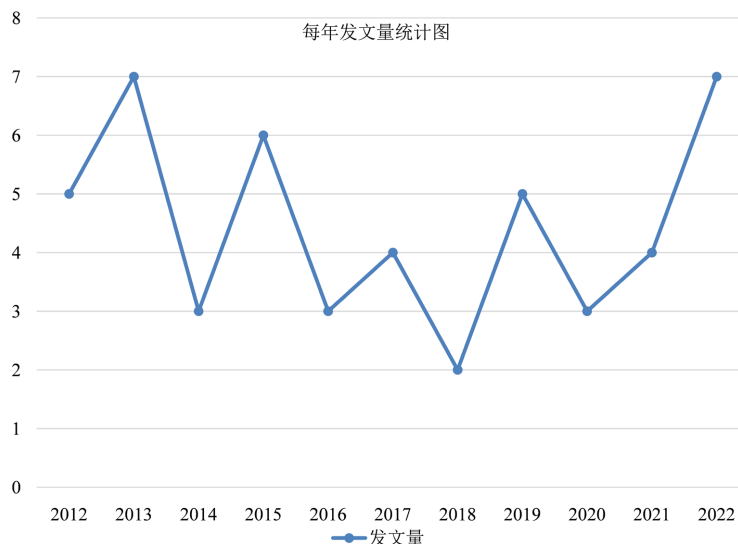


Figure 1. Statistical chart of the number of publications per year

图 1. 每年发文量统计图

2.1.2. 核心作者统计

2012~2022 年间, 在汉语中介语语料库建设研究方面, 专家张宝林、胡晓清、曹贤文、崔希亮等发文量较多。其中, 张宝林发文量在 10 篇以上, 排在首位, 其他学者的发文量均在 5 篇以下(见图 2), 可见汉语中介语语料库建设研究专家较为分散, 未能形成密切交流团队, 各学者之间的联系度需要进一步加强。

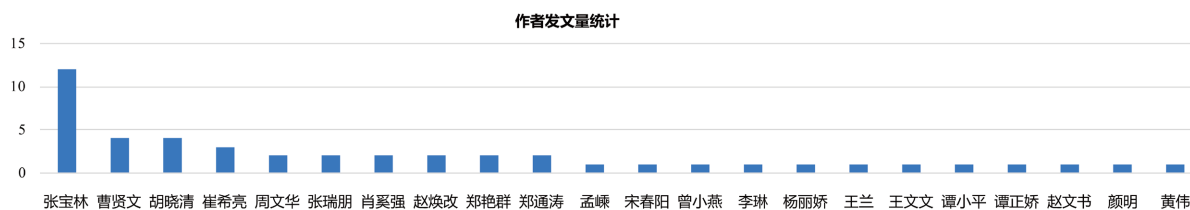


Figure 2. Core author statistics

图 2. 核心作者统计

2.1.3. 发文机构统计

机构的发文量可以在一定程度上体现出相关领域的研究实力及影响力。2012~2022 年间, 在汉语中介语语料库建设方面, 文献发表量前 7 名的机构是: 北京语言大学、鲁东大学、南京大学、南京师范大学、中山大学、厦门大学和西北师范大学。同时, 高校间存在互助合作, 南京师范大学与扬州大学共同发表——《论汉语中介语语料库建设基本问题》; 西北师范大学和福建师范大学合作发表——《关于汉语中介语语料库标注代码的思考》; 鲁东大学和韩国群山大学——共同发表《韩国汉语学习者中介语口语语料库的建设及意义》(见图 3)。可见汉语中介语语料库建设问题得到各高校的重视, 并且呈现资源互助和学科互补的发展趋势。

2.1.4. 期刊分布统计

2012~2022 年间, 有关汉语中介语语料库建设的 40 篇相关学术期刊发文分布如左图, 发文量前 7 名的期刊有: 语言文字应用(7 篇)、华文教学与研究(5 篇)、国际汉语教育(中英文)(4 篇)、世界汉语教学(4

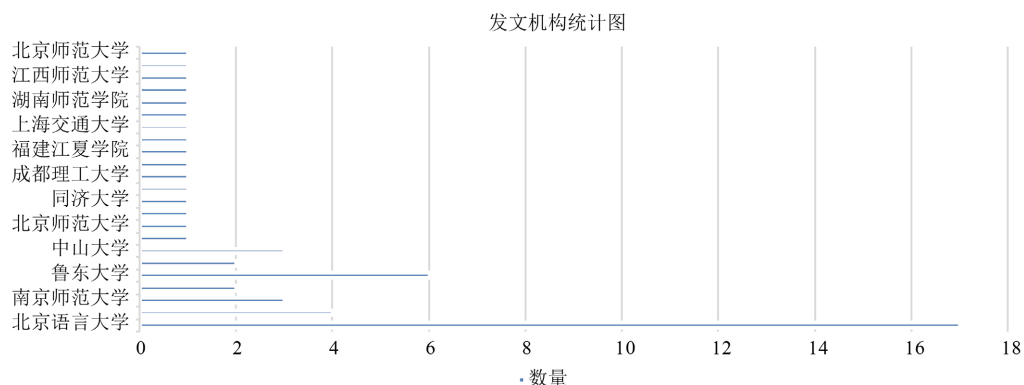


Figure 3. Statistics of issuing institutions
图 3. 发文机构统计

篇)、国际汉语教学研究(3 篇)、语言教学与研究(2 篇)、天津师范大学学报(2 篇)和海外华文教育(2 篇)(见图 4)。从以上数据可以看出,国内有较多期刊发表了一定数量的有关汉语中介语语料库建设的相关研究成果,反映出汉语中介语语料库建设这一话题备受关注,是语言学和教育学研究的话题。

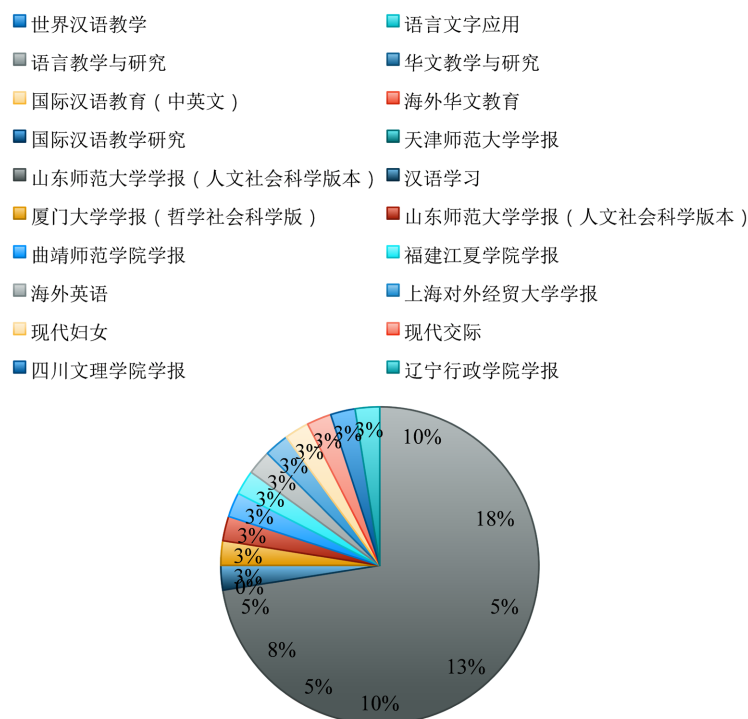


Figure 4. Periodical distribution statistics
图 4. 期刊分布统计图

2.2. 研究主题

2.2.1. 宏观层面

本文将收集到的 54 篇文献按宏观和微观两个层面进行研究主题分类。

在宏观层面涉及相关文献学术期刊 21 篇、学位论文 2 篇、学术会议 2 篇和中文图书 1 本,从总体设计、反思展望、研究热点和应用研究四个方面的分类如下:

1) 总体设计: 主要围绕汉语中介语语料库建设标准和建设基本问题。

围绕汉语中介语语料库的建设标准, 周文华从语料库内部和外部两方面提出要注意汉语中介语语料库建设的多样性和层次性[1]; 张宝林、崔希亮从建设流程、语料收集、语料标注、资源共享等角度谈汉语中介语语料库的建设标准[2]。

有关汉语中介语建设的基本问题, 张宝林从语料选取的原则、语料标注的内容与方法、语料的检索与呈现等方面谈汉语中介语语料库的建设问题[3]; 张瑞朋以中山大学汉字偏误中介语语料库为例探讨留学生汉语中介语语料库建设若干问题[4]; 颜明、肖奚强谈论了与汉语中介语语料库建设相互关联的有关语料性质和标注的四个基本问题[5]; 刘运同认为口语语料的选择、转写问题和转写清单的确立是汉语口语中介语语料库建设中的两个关键问题[6]。

2) 反思展望: 关于汉语中介语语料库的建设, 学者们注重阶段性反思并对未来发展提出展望。

21 世纪以来学者们密切关注汉语中介语语料库建设研究, 并进行及时的阶段性反思与总结。谭晓平对 2004~2013 年汉语语料库建设进行研究综述[7]; 蒋琴琴对 2010~2019 年国内汉语中介语语料库建设进行研究概述[8]; 王蕾等简评了 20 世纪后期北语一批标志性研究成果[9]。张宝林指出汉语中介语语料库自 21 世纪以来得到长足发展, 并由“简单而粗放”的 1.0 时代向“精细而丰富”的 2.0 时代转变[10]。

3) 研究热点: 依托先进科技进行汉语中介语语料库建库和科学研究

蔡武等基于 CiteSpace 软件对论文的发表年代、发文期刊、发文机构、核心作者及关键词聚类形成的研究热点, 并绘制可视化知识图谱, 提出汉语中介语语料库的数据来源存在局限性, 数据还不能充分共享; 同时还指出语料库建设缺乏跨学科视角, 不少技术问题仍有待突破[11]。谭正娇等从语料来源方面、建库方面和中介语检索的软件开发方面分析了目前国内的中介语语料库建设的特点[12]。

4) 应用研究: 中介语语料库的建设和研究与对外汉语教学、二语习得、偏误研究以及词汇对比等相互影响。

郑艳群分析了近些年来语料库技术在汉语教学中的代表性文献, 总结了语料库技术在对外汉语教学中应用现状[13]; 郝德才指出语料库语言学与对外汉语教学联系密切, 语料库技术服务于对外汉语教学[14]; 董婷婷基于韩国学生中介语语料库对词汇偏误进行研究, 并依据实验结果对韩汉词汇教学提出针对性教学建议[15]; 曲学智开发了现代汉语语料库和汉语中介语语料库的“汉语词语全句共现”的自动提取程序用以考察中介语和现代汉语之间的用法差异[16]。

2.2.2. 微观层面

在微观层面涉及相关文献学术期刊 25 篇、学位论文 7 篇、学术会议 2 篇, 本文从具体种类语料库建设、语料标注模式、语料来源和技术问题四个方面对文献的分类如下:

1) 具体种类语料库建设:

① 单模态、多模态、多维度汉语中介语语料库建设:

刘运同、胡晓清等关注口语中介语语料库建设和汉语学习者声调语料库[6] [17]; 在书面语汉语中介语语料库建设中, 玉兰关注留学生汉语错别字语料库建设[18]; 另外, 还有关于对外汉语虚词语料库的建设研究, HSK 动态作文语料库和中山大学汉字偏误语料库的相关应用分析。黄伟提出了汉语中介语多模态语料库建设方面的基本问题: 即在研究多维度语料库建设中以“全球汉语中介语语料库”的建设和研究为主[19]。胡晓清提出多维参照的汉语中介语语料库库群的建立构想[20]。

② 共时&历时汉语中介语语料库建设: 注重共识对比, 采集纵向发展语料、开展动态发展研究更是到学界关注。

何丹、张瑞朋比较研究北京语言大学 HSK 动态作文语料库、中山大学汉字偏误连续性中介语语料库, 对比三个汉语中介语语料库的建设和应用情况[21] [22]。曹贤文指出留学生汉语中介语纵向语料库建

设的若干问题[23]，赵焕改提出需求导向的汉语继承语学习者语料库建设研究[24]。

③ 通用&专门汉语中介语语料库建设：学界关注国别化汉语中介语语料库建设研究及应用分析。

有关国别化汉语中介语语料库建设：现有面向韩国的汉语学习者中介语口语语料库的建设和面向马达加斯加国的江西师范大学汉语中介语语料库建设。有关专门用途汉语中介语语料库建设：李代鹏提出建设对外汉语教材语料库[25]。

2) 语料标注模式：探索自动化标注方法，对语料标注的全面性存在争议。

张宝林提倡标注的全面性，对“偏误标注 + 基础标注”模式做了新的概括与解释；在标注内容方面，主张进行语义标注和语用标注，深化语篇标注和语体标注；在标注方法方面，倡导实行“有限的一错多标”；阐述了把可扩展的置标语言 XML 应用于“偏误标注 + 基础标注”模式[26]。与此同时，郑美平也倡导使用 XML 标注方法[27]。

而肖奚强、周文华对全面标注的主张持否定的态度，认为由于“支撑语义、语用、文体等标注的理论并不成熟”，“贪大求全并不可取，也不现实”，应“待相关的理论和实践研究比较成熟之后，再进行其他层面的标注”[28]。标注的类别应分为正确信息和偏误信息两类，标注赋码的类别应与标注内容的类别相匹配。

3) 语料来源：注重扩大语料规模和语料的平衡性问题。

李桂梅讨论了“全球汉语中介语语料库”在追求语料库平衡性方面应遵循的基本原则和采取的措施[29]；张宝林从国内、国外两方面提出了加以解决扩大汉语中介语语料库语料来源的方法[30]。

4) 技术问题：关注软件系统开发，助力语料库建设。

张宝林认为语言学专业人员应充分认识软件系统的重要地位与作用，积极参与研发工作，与软件技术人员密切合作，为语料检索提供高校、快捷的工具[31]。

2.3. 研究方法

通过统计文献中所使用研究方法的数量，可以看出有关汉语中介语语料库的建设的相关论文多采用文献分析法和定量与定性相结合的研究方法，而偏误分析方法和对比研究方法在研究中使用较少。(见图 5)其中计算机辅助法在研究中使用普遍，这也说明有关汉语中介语语料库建设研究更加注重数据的科学性和善于利用科学技术手段更新研究思路和方法。

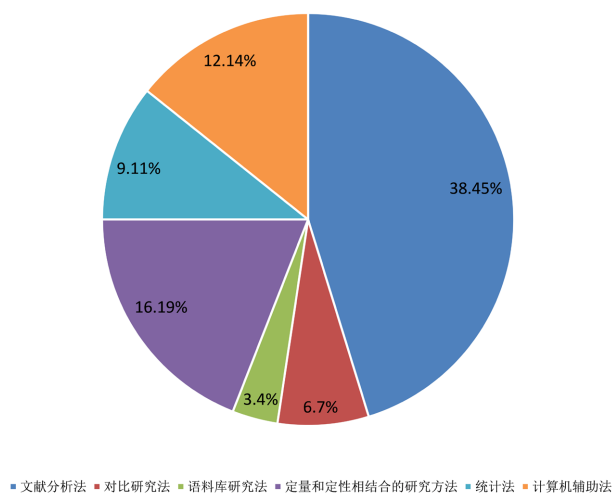


Figure 5. Statistical diagram of research methods

图 5. 研究方法统计图

3. 研究不足

针对以上数据的分析,下面将从文献发表不足、研究主题不足和研究方法不足这三个方面对国内近十年汉语中介语语料库建设研究存在的问题进行分析。

3.1. 文献发表不足

(1) 从发刊机构来看,高校间存在学术资源互助,但这种现象并不普遍。这也侧面说明了在汉语中介语语料库近 30 年的建设历程里,始终存在着资源共享的问题。绝大多数由国家投资建立的语料库建成之后只供本单位使用并未对外开放。在现已建成的语料库当中,只有 HSK 动态作文语料库、中山大学汉字偏误标注连续性中介语料库、暨南大学留学生汉语中介语料库和全球中介语料库四所语料库对外开放,语料库的封闭造成了学术资源和社会效益的浪费。

3.2. 研究主题不足

(2) 对语料标注的广度深度的研究不够,没有一套明确统一的语料标注编码。在标注同一语料时,不同汉语中介语语料库采用的标准具有较大差异,可见明确统一的语料标注通用符号和规则是汉语中介语语料库发展的有效突破口。同时,关于标注是否全面的问题还有待商榷,需要依据现实的建库需求调整 and 适应,在建库之初就应明确。

(3) 缺少有关扩大语料规模方式的研究。国内现存的中介语语料库规模普遍较小。虽然全球中介语料库的语料总规模已经达到了 1.26 亿字,是中介语料库 2.0 时代的标志性成果,侧面反映了语料库规模在逐渐扩大的趋势。

语料库规模小的原因主要有以下两点:一是语料收集难度大。在语料收集过程中,首先要征得被试者同意,其次,要进行个案的描述和分析,在动态追踪方面,很难一直坚持下去,在此期间,还要克服双方交流壁垒,最后经过对案例进行随机抽样筛选后,样本的数量层层递减导致语料库规模小。二是语料收集方式单一且仅只进行一维层次研究,从书面语语料收集来看,大多采用收作业的方式,口语语料库的语料采集大多采用面对面录音转录的方式,因而受时空限制很大,影响样本收集数量。

(4) 缺少共时汉语中介语语料库和汉语母语语料库对比研究、汉语中介语语料库和国内英语学习者语料库对比研究。共时研究和历时研究的不平衡性根源于语料收集的不平衡性。这里的不平衡性表现在以下三个方面:一是语料收集类型的不平衡性,在中介语语料库建设过程中,书面语语料收集量明显大于口语语料的收集量;二是语料收集对象水平的不平衡,在对初级汉语学习者的语料收集较少,中高级汉语学习者的语料收集相对较多;最后是语料收集的国别具有不平衡性,在外国留学生当中,来自欧美国家汉语留学生学习语料相比较其他国别的留学生来说收集得较少。

(5) 缺少语料分析工具研究,关于自动分词和语音转录软件有待开发。有关标注模式,现在大多采用人工校对,自动标注的质量不能达标,技术关口需要相关学科的带动。同时还要注意提升人工的标注素养,在人工标注方面,标注人员需得接受语言专业技能培训,以确保校验结果的准确性。

3.3. 研究方法不足

(6) 研究方法单一简单,以文献分析法为主,缺少数据分析研究支持。计算机辅助法在研究中使用需要重视,计算机技术在研究中的应用可以大大提高数据分析的准确性和科学性。在汉语中介语语料库建设过程中,有关扩大语料收集规模、提升自动校对的准确度和研究分析的科学性都需要相关技术软件研发来助力语料库建设研究。同时提倡多元研究方法的使用,在保证基础数据真实、准确的基础之上,多样化的研究方法会大大提高研究的准确性、科学性和可信性。

4. 研究展望

4.1. 文献发表

(1) 增加跨学科合作研究,注重高校间资源共享。利用高校间不同优势,实现学术资源互助和优势互补。例如以课题为导向,一方提供语料资源,另一方提供语料分析技术支撑,双方合力构建特色语料库;实现学科互助,合力提升跨学科技能。例如语言学学科与信息技术跨学科进行中文信息处理,从而达到高效建库,数字化分析。

(2) 增加学者之间的互动交流,促进观点碰撞。学术会议的定期举办可以为学者提供观点交流的平台,有效地推动问题的发现和解决。例如:定期举办线上或线下的会议,通过会议平台,加强学者间交流互动的同时,也有助于研究团体的形成和打造。

4.2. 研究主题

(3) 加强语料标注广度和深度的研究,确立一套统一的语料标注代码,为语料规范化标注提供依靠。确立一套业界统一共识的语料标注符号系统是汉语中介语语料库建设亟需解决的问题,以及在有关标注标准方面进行讨论并达成共识,其次,还要借助科技研发手段,创新语料标注和收集的形式,以期实现高质量语料标注。注重语料收集方式的多元化,创新收集方式,扩大语料来源途径。例如,可以利用处在一线工作的国际中文教师,保障语料的新鲜真实,以及语料收集工作的可持续性。

(4) 加强国别化语料库和追踪语料库的建设研究。首先,随着汉语作为第二语言研究和学习的不断深入,建设面向不同国别的具体中介语语料库是未来发展的方向;另外在收集语料时,要注意语料的完整性和发展性、共识性和历时性。例如:对语料收集对象进行动态追踪、规范收集对象信息等,这些可以为动态追踪语料库的建设提供基础。

(5) 增加有关自动分词、语音转写软件的开发研究。计算机技术的加持会大大提高汉语中介语语料库语料收集和数据分析的效率。例如:语音转写软件上线后,将进一步提高语音转写的效率,从而降低口语语料收集难度,增加口语语料来源,助力汉语口语语料库的建设;同时自动分词技术可以减轻人工标注的压力,为标注提供相对科学的分析语料。

4.3. 研究方法

(6) 增加有关共时汉语中介语语料库和汉语母语语料库、汉语中介语语料库和国内英语学习者语料库的对比研究。通过国内外共时中介语语料库对比,挖掘不同中介语学习的差异与共性,借鉴国外中介语语料库建设优势,学习其先进理论和研究方法这可以为国内汉语中介语语料库建设和研究提供新发现和新办法。

5. 结语

根据上文对 CNKI(中国知网)数据库中近十年来收录期刊文献的研究,尤其是通过从文献发表情况、研究主题和研究方法等方面对国内汉语中介语语料库建设相关研究成果的分析,得出的分析结果对国内汉语中介语语料库建设和研究带来一些启示。首先从文献发表情况来看,应该更加关注汉语中介语语料库的共享问题,重视学科间的互助交流和高校间的资源共享;同时,从还要提供学者交流互动平台,助力研究团队的形成。其次,从研究主题来看,汉语中介语语料库语料来源和平衡性问题值得持续关注。创新语料收集方式,探索扩大语料来源途径,可以通过一线国际中文教师获得新鲜可持续性的语料。同时,还要加强语料标注广度和深度的研究,明确一套统一的语料标注代码,是汉语中介语语料库建设目前亟待解决的问题。另外,重视技术探索,增加有关自动分词、语音转写与软件的开发研究将会大大提高

语料收集和数据分析的效率,减轻人工标注的压力。最后,从研究方法来看,对于汉语中介语语料库建设的研究应更加注重研究方法的多元化,注重使用对比研究法,加强共时汉语中介语语料库和汉语母语语料库、汉语中介语语料库与国内英语学习者语料库对比研究。

总之,汉语中介语语料库建设,经过了30年的风风雨雨,在21世纪进入蓬勃发展期,汉语中介语语料库日益成为业界研究热点并不断实现自我突破,同时与国际中文教育事业紧密联系,相辅相成。虽然目前暴露出有关语料收集、标注以及资源共享等方面的问题,但是我相信在未来,中国将会建成更大,更全,更方便的实时共享语料库平台,并为其他学科的发展提供新的动能!

参考文献

- [1] 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(6): 97-105.
- [2] 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(2): 125-134.
<https://doi.org/10.16499/j.cnki.1003-5397.2015.02.017>
- [3] 张宝林. 关于汉语中介语语料库建设的若干重要问题[C]//中文教学现代化学会. 数字化汉语教学: 2012. 北京: 清华大学出版社, 2012: 325-331.
- [4] 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(3): 133-140.
<https://doi.org/10.16499/j.cnki.1003-5397.2013.03.022>
- [5] 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(1): 136-144.
<https://doi.org/10.16499/j.cnki.1003-5397.2017.01.015>
- [6] 刘运同. 汉语口语中介语语料库建设中的两个关键问题[J]. 华文教学与研究, 2020(1): 47-52.
<https://doi.org/10.16131/j.cnki.cn44-1669/g4.2020.01.008>
- [7] 谭晓平. 近十年汉语语料库建设研究综述[C]//北京大学对外汉语教育学院. 第七届北京地区对外汉语教学研究生论坛文集. 北京: 北京师范大学中文信息处理研究所, 2014: 26-31.
- [8] 蒋琴琴. 近十年国内汉语中介语语料库建设研究概述[J]. 海外英语, 2019(3): 82-83.
- [9] 王蕾, 赵琪凤, 蒋荣. 简评20世纪后期北语一批标志性研究成果[J]. 国际汉语教学研究, 2022(3): 37-44.
- [10] 张宝林. 从1.0到2.0——汉语中介语语料库的建设与发展[J]. 国际汉语教学研究, 2019(4): 84-95.
- [11] 蔡武, 郑通涛. 我国汉语中介语语料库研究现状与热点透视——基于CiteSpace的可视化分析[J]. 华文教学与研究, 2017(3): 79-87. <https://doi.org/10.16131/j.cnki.cn44-1669/g4.2017.03.011>
- [12] 谭正娇, 王文文, 余晓铃. 汉语中介语语料库研究热点分析——兼谈对国际中文专业的影响[J]. 现代交际, 2021(16): 75-77.
- [13] 郑艳群. 语料库技术在汉语教学中的应用透视[J]. 语言文字应用, 2013(1): 131-138.
<https://doi.org/10.16499/j.cnki.1003-5397.2013.01.026>
- [14] 郗德才. 基于语料库研究的对外汉语教学[J]. 辽宁行政学院学报, 2012, 14(11): 106-108.
- [15] 董婷婷. 基于“韩国留学生汉语中介语语料库”的词汇偏误研究及应用[D]: [硕士学位论文]. 烟台: 鲁东大学, 2012.
- [16] 曲学智. 基于全句内共现的现代汉语和汉语中介语词汇知识对比研究[D]: [硕士学位论文]. 北京: 北京语言大学, 2017.
- [17] 胡晓清, 许小星. 韩国汉语学习者中介语口语语料库的建设及意义[J]. 华文教学与研究, 2020(1): 53-59.
<https://doi.org/10.16131/j.cnki.cn44-1669/g4.2020.01.009>
- [18] 王兰. 留学生汉语错别字语料库的建立与研究——以昆明理工大学为例[J]. 现代妇女(下旬), 2014(12): 5-6, 8.
- [19] 黄伟. 多模态汉语中介语语料库建设刍议[J]. 国际汉语教学研究, 2015(3): 60-66.
- [20] 胡晓清. 多维参照的汉语中介语语料库库群的建立构想[C]//中文教学现代化学会. 第十届中文教学现代化国际研讨会论文集. 北京: 清华大学出版社, 2016: 384-389.
- [21] 何丹. 三个汉语中介语语料库对比分析[D]: [硕士学位论文]. 广州: 中山大学, 2012.
- [22] 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(2): 131-136. <https://doi.org/10.16499/j.cnki.1003-5397.2012.02.031>

-
- [23] 曹贤文. 留学生汉语中介语纵向语料库建设的若干问题[J]. 语言文字应用, 2013(2): 127-134.
<https://doi.org/10.16499/j.cnki.1003-5397.2013.02.018>
- [24] 赵焕改. 需求导向的汉语继承语学习者语料库建设研究[J]. 国际中文教育(中英文), 2022, 7(2): 15-22.
- [25] 李代鹏. 基于数据挖掘技术的汉语教学语料库建构[J]. 海外华文教育, 2019(1): 115-124.
<https://doi.org/10.14095/j.cnki.oce.2019.01.014>
- [26] 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(5): 27-34.
- [27] 郑美平. 汉语中介语语料库 XML 标注方法的应用——基于 HSK 动态作文语料库的分析[J]. 福建江夏学院学报, 2015, 5(6): 99-105.
- [28] 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(3): 368-377.
<https://doi.org/10.13724/j.cnki.ctiw.2014.03.007>
- [29] 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(2): 46-51.
<https://doi.org/10.16131/j.cnki.cn44-1669/g4.2017.02.003>
- [30] 张宝林. 扩大汉语中介语语料库语料来源的途径[J]. 国际中文教育(中英文), 2022, 7(2): 30-37.
- [31] 张宝林. 汉语中介语语料库检索系统透视[J]. 天津师范大学学报(社会科学版), 2021(6): 29-37.