

国际中文教育专硕生利用语料库进行学位论文写作的调查与分析

王昌宇

刘运同 *

同济大学, 中国

摘要

语料库已成为国际中文教育领域撰写毕业论文的常用方法或工具之一。本文基于知网、读秀等学术平台的搜索, 对 2010-2023 年国际中文教育专业 (含改名前的“汉语国际教育”“对外汉语教学”) 专硕生的论文进行定量考察, 聚焦目前使用语料库法进行学位论文写作的选题和研究方法。利用 Tableau、CiteSpace 等数据分析软件, 在定量分析和定性分析的基础上, 描述分析目前各类选题范围的研究现状, 并明确亟需注意的问题与不足, 最后提出相应改进意见和建议。

关键词

语料库, 专业硕士学位论文, 选题, 方法

1 综述

近年来, 随着语料库的大规模建立, 语料库研究方法已广泛被应用于国际中文教育研究的各个领域。在国际中文教育专业硕士论文写作中, 也已经有了一定的数量。专业硕士论文的选题与方法, 体现出硕士群体的学术关注重点、学科素养与专业能力, 因此, 专硕生学位论文的选题一直受学界的关注和重视。黄伟、李珺婷 (2022) 曾对 20 年间 (2001-2020) 的偏误类硕士学位论文进行了计量分析, 指出当前文章在偏误类型和偏误原因方面囿于成说, 缺乏具体问题具体分析合理性; 王辉 (2022) 对 2009-2020 年间汉语国际教育专业硕士论文被引情况进行了分析, 指出当前的研究热点和研究趋势; 李丽 (2022) 对文化类选题进行过仔细分析, 并结合困境给出相应对策; 黄微 (2022) 对专业硕士文化类选题进行了较为全面的分析, 指出论文类型单一, 缺乏创新等问题。也有人对各高校的国际中文教育专硕论文做过专题调研, 如徐晶晶 (2022) 《近十年甘肃高校汉语国际教育硕士学位论文选题研究》等。

但是截至目前, 在国际中文教育领域内, 还没有一篇围绕使用语料库方法进行学位论文写作的全面综述类文章。本文以近十余年国际中文教育专业硕士蓬勃发展时期的学位论文为基础考察对象, 围绕选题与方法两个维度展开, 描述分析各类学位论文的特点。分析不足的同时, 给出相应建议。

2 选题的定量统计与分析

结合知网、读秀等学术平台，依据文章标题、摘要及关键词对各高校论文进行高级检索，可得近十余年使用语料库进行学位论文写作的总计 359 篇。其中部分大学硕士论文因未上传知网等学术平台等原因，未能搜索到相关文章。但缺失的文章数量因所占比重不大，对研究结果的影响甚微。

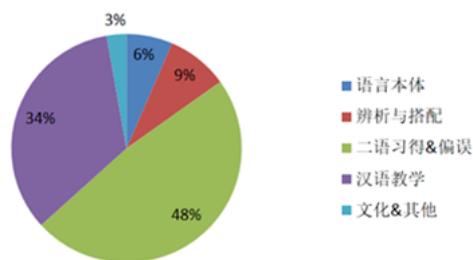
从历年的论文数量上来看（图 1），2011 到 2023 年总体呈上升趋势，其中 2021 年达到峰值，有 53 篇之多。2022 年略有回落，2023 年仅有 10 篇，可能与一些大学停招专硕士生有关。

图 1. 使用语料库的专硕论文数量



就 359 篇文章的选题来看，内容虽然广泛，但主题相对集中。涉及的领域大致有汉语教学语言本体、汉语教学教师专业知识、习得过程中的偏误以及文化等其他内容。因此，根据文章内容及讨论方便，我们将上述文章分成了 5 个大类，分别是：语言本体类、二语习得与偏误类、辨析与搭配类、汉语教学类、文化与其他类，其中语言本体类是指围绕语音、词汇、语法等任一或多个角度展开专题论述的文章，辨析搭配类是指利用语料库对一组同义词进行辨析，辨析搭配类的文章数量较多，且与纯粹的本体研究有所不同，故而单独成类。各部分占比如图 2 所示。

图 2. 使用语料库进行学位论文写作的各类文章占比



在 5 种类别的文章中，二语习得与偏误类占比最多，共计 159 篇，占 44%。文化与其他类占比最少，总计 9 篇，占比 3%，该类中的其他主要是指一些综述类的文章，如《汉硕学位论文“偏误研究”类选题分析》、《黑龙江大学汉语国际教育硕士学位论文考察》，也包含一些难以归入他类的当下热点文章，比如针对国别化语料库展开的调研——《泰国汉语教学语料库现状分析》等。

2.1 语言本体类选题

在语言本体类选题中,语言对比是一个比较热门的选题视角。李玉楠将汉语和英语中的爱恨类动词进行了比较;顾爽将汉英常用的贬义词进行对比研究;再如潘静静将二语学习者母语和汉语的假设标记“的话”的使用情况进行对比研究。上述提及的这些论文虽然是本体研究,有些聚焦于语言间语法的不同,有些聚焦于语义和语用,本质上都是立足于汉语教学,对于二语教学都有直接的启示作用。

图 3. 语言本体选题占比



由于近几年一直强调以实践性为导向,所以专硕生的学位论文也随之出现了转向,偏实践性的论文越来越多,但这也导致当前的语言本体类选题数量急剧下降。事实上,面向教学展开的语言本体研究是很有启发意义的,这样的研究选题应该继续鼓励。

就选题内容看,当前语言本体研究存在比例严重失衡的问题,其中最为突出的,是使用语料库进行语音相关研究的数量为0篇。语音方面的缺失,究其根本,笔者认为有两个主要原因:一是汉语口语语料库的开发与建设还处于起步阶段,虽然现在已经有建成且开放的不少大规模的口语语料库,但其数据本身还需要校对,推广至专硕生群体也还需一段时间;另一个原因是涉及语音学的口语语料处理需要一定的相关专业知识,国际中文教育专硕生因其专业属性和培养目标不同,而语音处理的相关知识和软件操作能力的培养也需要一段不短的时间,所以国际中文教育专硕生在语音数据的处理能力上也略显薄弱。

语言本体中其他子类研究中也存在一些有待进一步完善的地方。词汇选题上当前谓词和语气词研究较多,比如鄢芹芹(2021)围绕“吃”一类的有关动词展开论述、石芳联(2021)对动态助词“了”展开了历时考察等。名词的研究几乎集中于汉外对比,其他词性的词汇研究非常稀缺,可以适当增加该类选题。语法上除了陈如岳(2016)的一篇动词结构研究,其他语法研究均是关于话语标记,有如“X的是”、“的话”等研究。

2.2 二语习得与偏误类选题

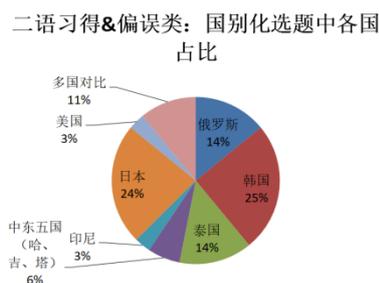
利用语料库进行二语习得与偏误研究的研究者一直居高不下,语料库因包含大量留学生使用汉语进行交际或写作时存在的真实错误,故而围绕偏误而展开的研究类别也非常之多。其中对因国别和汉语水平差异而产生的偏误研究最多。

图 4. 二语习得 & 偏误类选题数量占比



在统计的近十余年专硕论文中，国别化的偏误研究共计 66 篇。其中，汉字文化圈国家占比最大（韩国 17 篇，日本 15 篇），共占据该小类研究的国别化研究的 48%。至于多国对比的选题，大多是基于自己所教授的班级情况，如混班制等，这类偏误研究数量也不算少且具有现实启发意义。

图 5. 国别化选题中各国占比



我们对现有国别化选题中最主要的四个国家选题做张一览表，便于更直观地看出选题研究的分布情况。

表 1. 国别化选题中研究主题分布情况

	日本	泰国	韩国	俄罗斯
词汇	方位词（上）； 介词（在，跟）； 疑问语气词； 语气词（吧、吗、呢）	方位词； 动态助词（着）； 频率副词；	汉语介词； 动态助词（着）； 趋向动词（X 来）； 否定词（别）； 疑问代词（怎么）； 人称代词（自己）； 副词（真、太、很）；概数词；	程度副词（2 篇）； 指示代词； 离合词
句式	定语习得； 副词+是； 句式（是...的）； 汉语能性结构； 动宾语序	多项定语； 因果复句； 反问句； 可能补语； 情态补语； 句式（在...上）	“有”字句习得；比较句； 时量补语； 动宾结构； 形容词谓语句	语序； 汉语宾语
汉字	汉字习得（2 篇）；	/	汉字书写情况； 形声字习得	汉字偏误（中 1 篇、外 1 篇）
写作	书信写作偏误	语篇衔接	书信写作； 写作语序（2 篇）；求职信	/

从上表我们不难发现,无论是从汉语要素还是汉语技能教学角度,结合语料库研究进行的选题都主要集中于学习者在学习汉语时可能遇到的难点:疑问语气词、多项定语、语序、汉字习得、书信写作等等。这其中值得我们注意的是:因部分开设国际中文教育专硕的学校,会同时招收部分外国学生进入该专业学习,所以针对同一个重难点的偏误情况,中外学生不同的研究视角和结论是值得重视的。以汉字偏误为例,在《俄罗斯留学生汉字习得偏误分析》一文中,作者基于自己国家(俄罗斯)学生的汉字书写情况,在考察偏误时,细致地将学生按汉语水平分成了A、B、C三类,并根据同一汉语水平下的汉字书写情况作了对比考察。而同校同类型的两篇文章也是基于国别化而做的文章,虽然在数据选择上更为充实,且对偏误类型和偏误原因作了更深入地探讨,但是却未能对学生汉语水平进行进一步细分。

在依据汉语水平而做的偏误研究中,我们应该关注到,针对中高级的留学生偏误研究有14篇,而围绕初级留学生汉语水平的论文仅1篇。很显然,针对初级汉语水平学习者而做的研究严重不足,应给予更多的关注度。

表 2. 基于汉语水平而做的偏误研究数量情况

	数量	例文
初级	1 篇	王瑞雪《对外汉语初级阶段的常用程度副词偏误研究》
中高级	14 篇	薛楚楚《基于 HSK 语料库的中高级阶段留学生关联词语偏误分析与教学》

鲁健骥(1992)曾提出,“虽然课堂上的学生可能来自于各个大洲各个不同的国家,但是偏误就形式上而言,不外乎四大类型——遗漏(少成分)、增添(多成分)、替代(所用不当)、错序(词序有误)。”同时,教学上的失误是诱发的学生偏误的原因之一;二语习得与偏误类研究方面,预测偏误、找出偏误固然重要,但能否恰当地运用于教学,对学生汉语水平有切实的提高,是更为重要的问题。另外一个很重要的问题是,目前的偏误研究大多着眼于学习者的母语迁移,也就是受母语影响而造成的偏误,但实际上,偏误还可能与教师的教学方式、授课时的表述有着密切关系,而当前这部分研究是严重缺失的,所以在二语习得与偏误研究中,我们不妨可以通过对老师授课的实际观察和长期记录,来探索基于教师教授的方法、表达,所导致的教授班级学生偏误的情况,这也有利于提高教师教学能力和科研能力。

2.3 辨析与搭配类选题

辨析与搭配一直是专硕生利用语料库进行研究的一个长期关注话题,每年几乎都有数篇。从占比来看,2017年之前大约在3%左右浮动,2018年达到峰值,占比24%,此后几年都在6%~16%区间浮动。在总计28篇的该类研究中,有关词汇的辨析与搭配文章最多,有25篇。其他该类文章主题包含1篇句式的辨析,以及2篇同音形近义词研究。在词汇辨析中,有些是单纯地将两个或多个词语进行对比,并探讨其在教学中应该注意的问题,该类文章总计17篇;有些是面向某个具体的国家而进行的国别化词汇辨析,总计有8篇。

辨析与搭配类选题大部分集中在词汇,而从2011-2023年,12年间句式的辨析仅有1篇。外国留学生在二语习得的过程中,句式的偏误有时候不仅仅是单一句式习得问题的体现,有时候是受相似句式的影响而造成的。对句式偏误的研究虽然有难度,但仍是值得研究的重要方向。另一个不足体现在同音形近字的研究上,汉语因为有着大量的同音字,这对于外国学生学习来说,是一个不小的困难。这方面的研究也可以进一步加强。

图 6. 辨析与搭配类选题数量占比



2.4 汉语教学类选题

汉语教学类选题是利用语料库进行学位论文写作的一个主要类别，就数量上看总体呈现上升趋势，结合近十余年百分值占比，该类选题每年平均稳定在 30% 左右，而 2023 年的占比更是增长至 70%。在归类时笔者将教材、教法和教师（“三教”问题）研究都归入了汉语教学这一类别之中。其中与教学（法）相关的共计 97 篇，和教材有关的有 15 篇，仅 1 篇选题和教师相关。

我们使用 Tableau 对上述 103 篇文章进行主题高频词分析，可得图 6 的主题词透视数据图。左侧与词语配对的是作为选题出现的次数，右侧表格的方格大小表示该主题词在该类选题中的占比和重要程度，这其中占比最高的是句式和词汇，也就是说针对不同句式和词汇而进行的教学设计和研究最多，这也符合实际教学需要。上述统计结果基本符合事实。但是考虑到软件无法结合文章实际内容等问题，我们采用人工方式重新统计了各部分的数量。

图 7. 汉语教学类选题数量占比

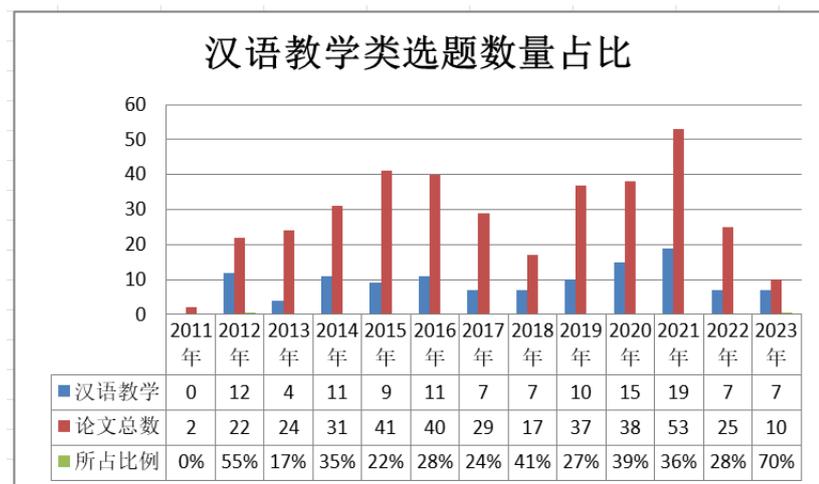


图 8. 汉语教学类——主题词数据透视表

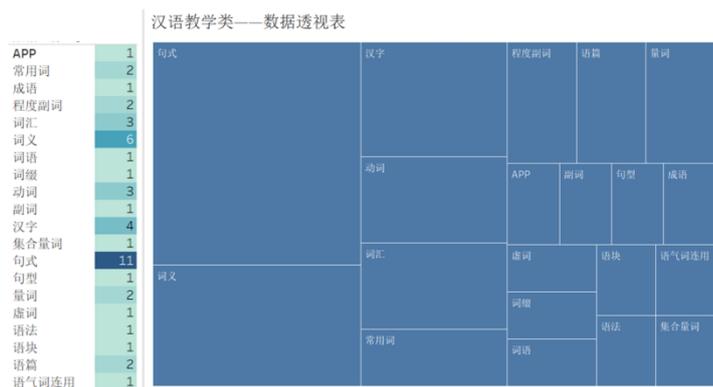


表 3. 汉语教学研究分类数据表

基于实践的研究	句式 & 语法	22 篇	词汇教学	54 篇
	汉字教学	4 篇	语用	10 篇
	文化教学	3 篇		
基于理论的研究	教学法 & 理论		3 篇	
其他	学习资源 & App 探究		3 篇	

通过人工统计，我们重新整理相关数据，可以发现，目前大部分研究都是基于实际教学的语言点而进行的，只有 2 篇涉及了相关教学法和教学理论。然而理论指导下的教学研究是非常重要的，当前语言学界有很多语言教学理论，早期的传统语法、生成转换语法，现如今流行的构式语法、语块理论等。之所以提倡在相关理论下进行语言教学，是因为利用教学理论去进行教学，有利于我们更加深刻地认识理论本身，充分利用其优点，增加对语言的认识。教材研究中，就单本教材展开的专题调研和就某一主题展开的多本教材探究数量平分秋色，各占 9 篇。单本教材研究中围绕《博雅汉语》展开的有 1 篇，围绕《发展汉语》展开的研究有 6 篇。综合教材的研究主要集中于词汇或句式的编排，也就是探讨某一语言点的教学具体路径，也有一些是针对词汇对比而进行的。教材作为语言知识的载体，其研究的重要性仅次于教学，如何编排、遵循什么样的顺序、语言点之间如何搭配和展现是该类研究中非常重要的问题。但是综观当前的研究，围绕某一知识点的纵向研究很多，但是不同教材之间的横向对比相对较少，可以适当增加类似文章。

表 4. 汉语教材研究主题 - 数量表

教材名称 / 类型	研究主题	数量
《发展汉语》	教材语体词、生词的英文释义、 时间副词、成语、双音羨余词、 否定是非问句	6 篇
《博雅汉语》	汉字教学	1 篇
综合教材	编排研究 (“被”字句 / “是” 字句 / 连动句 / 形容词编排生动 形式)、词汇对比	5 篇
教材中的文化因素	国家形象	1 篇

从数量上看,使用语料库进行的汉语教学研究中,有关教师的研究非常缺乏,目前唯一的1篇论文是关于教师行业用语泛化研究。该类研究的缺乏倒是合乎逻辑,因为结合创新性的要求,该类选题确实较难找到切入口。教师作为教学活动的主体,其研究的重要性是无可厚非的,但是如何结合语料库进行研究,参考已刊论文也许可以适当拓宽思路,比如针对教师在教学中自行建立小型语料库的方法、提高使用语料库进行教学和科研的能力等,更多的选题有待在实际教学中做更深入地探索。

2.5 文化与其他类选题

在语料库法的研究视野下,文化与其他类选题是比较少的一类。在为数不多的论文中,主要涉及的主题有:文化教学策略、文化内涵分析、文化认同和国家形象等。

图9. 文化与其他类选题数量占比



从上面的数据,我们不难发现,专硕论文中的文化与其他类选题目前鲜有人涉足,利用语料库进行文化教学及相关研究的,更是屈指可数。这其中一个很重要的原因是很多学生想不到语料库可以进行文化研究。其实语言中有很多和文化相贴合的部分,比如词义与文化,祖晓梅在《跨文化交际》一书中就曾将词义与文化分为四个部分来论述——颜色词、动物词、禁忌语和委婉语。高樱2017年所写的《“青”的文化内涵及对外汉语教学对策研究》一文就属于颜色词的研究。这样类似的话题其实还有很多,如有着不同文化含义的“猫”在汉英等双语平行语料库中,能够检索出“长舌妇”、“包藏祸心的女子”等类似的语义,而这样的语义在中文中是没有的,而这正是语义在文化中的体现,会影响人们的日常交际。所以利用语料库来进行文化研究是一个可行的方向,但是需要有特定的依托,最为通常的研究视角是基于跨文化视野下的语言交流活动。

通过对五类选题进行定量分析,我们不难发现每个类别的文章中目前都存在一些不足与问题。结合各小类的问题,我们可以对专硕生利用语料库进行写作中选题的不足做一个总结。周小兵(2017)等学者指出汉语国际教育专业学位论文的研究材料必须真实可靠,研究应当用于解决汉语国际教育中的实际问题。通过对以上选题的定量分析,我们可以看到语料库在国际中文教育专业硕士毕业论文写作中,目前存在着三大不足:一是在使用语料库进行专业硕士学位论文写作时,涉及语音等有一定难度的选题,学生鲜有涉及;二是大多学生已经习惯于把语料库当成搜集典型偏误例证的来源,而非发现学生错误的数据来源;三是选题时未能充分结合外

国留学生学习的重难点,囿于现有研究框架,生搬硬套现象明显。针对以上的不足我们给出如下三点总结性建议:第一,选什么样的题目去做,落实到每个学生,其实都是个体的选择,但是作为学生而言,通常会选择数据易得、操作不太复杂的选题。因此作为论文指导教师,应该鼓励一部分有能力的学生挑战较难的选题,以打破目前部分选题无人问津的现状;第二,现有各种各样的语料库,可以进行多种类型的研究。较为集中的偏误与习得研究习惯于套用母语负迁移等习得偏误理论,把语料库的数据看成是验证现有理论的例证,而非通过学生的偏误来探寻学生错误形成的原因。针对这样的现象我们倡导在教学实践中,结合偏误的共性来完成个性化的调研,也就是允许学生结合已有理论,但同时必须总结出自己所教授的学生群体的个性问题;第三,中国学生因为不能切身体会外国留学生学习汉语过程中的真实困难,所以在论文中未能充分结合学习重难点而进行探讨也很容易理解,解决这一问题也并不困难。当前国际中文教育专硕生群体中有一部分是在华学习的外国留学生,他们的专硕论文对中国学生来说有一定的启发意义,中国学生可以结合他们在选题时的切入视角,从母语为汉语者的角度做更深的解读。

3 使用语料库进行学位论文写作所选用方法的综述

研究方法方面我们重点关注的是各小类论文语料的选取和来源,以及使用了什么样的方法去论证文章主题。

在语言本体类研究中,汉语的语料几乎都来自北京大学 CCL 语料库和北京语言大学 BCC 语料库或是与自建语料库相结合的方式使用。与此同时,基于同一语言点的对比是该类占比最多的文章类型,因此就语料选择而言,平行/多语语料库成为了研究者的首选。根据研究目的的不同,当然也有一些其他语料来源,比如在研究词汇感情色彩时,部分研究者从《汉语褒贬义词语用法词典》、《常用贬义词语讲解辞典》等辞书中选取了对应语料,就考试而言所做的研究通常选用《汉语词汇与汉字等级大纲》作为语料。在本体研究中,当前用于解释的方法大多也是基于三个平面理论,少数几篇仅从语用或某个单一角度进行论述。从方法所结合的理论来看,目前在本体类研究中,基于动词而做的传统语法和生成语法仍占主要地位。

二语习得与偏误类研究论文中,语料选取来源最多的是 HSK 动态作文语料库,在 159 篇论文中,使用频次达 149 次。而近 5 年汉字偏误的热度持续走高,因此在做汉字偏误时,除 HSK 动态作文语料库外,中山大学的汉字偏误标注的汉语连续性中介语语料库是被使用频次最高的语料来源。该类文章的论证方法也几乎趋于统一,选取研究所需的相应语料,进行统计分析,对所发生的偏误进行分类,结合各类偏误类型总结偏误原因。针对国别化和汉语水平而做的研究,除语料选取外,论证思路几乎无差别。

辨析与搭配类文章的目的是为了通过比较,找出词汇、句式或汉字中的相同与差异,在分析学生易发生偏误的基础上,针对偏误的特点对差异进行有针对性的讲解和教学。这样的研究目的也直接影响了我们研究方法的选择。语料的选取大多是来源于语料库,其中也有一些是利用调查问卷来获得语料,使用调查问卷作为语料来源通常有特殊原因,一般是出现在面向国别化的词汇辨析中。语料库选取上,北京语言大学的 HSK 动态作文语料库是被使用频次最多的语料库。大部分学生在利用语料库检索过程中通常会同时利用多个语料库进行分析。研究方法上,主要体现在三个层面:一是如何描述两个词汇或句法结构间的差异,二是如何描述多个/一类词汇或句法间的差异,三是针对差异如何进行有效教学。

解释差异并有针对性的进行有效教学,是汉语教师的一项核心能力,在解释词汇差异时,最常用的解释维度有三个——句法、语义、语用。汉语三个平面的理论用于充当解释词汇差异

的三个维度既便于操作，又具有解释力。在超9成的论文写作者都使用该方法去解释和描述差异的情况下，有一些方法的使用便会使人耳目一新。比如有学生只针对一组词间的交叉义项进行辨析，再通过 WordSmith 计算出误代率，从而归纳二者的误代原因。这样的创新做法是值得提倡的。

汉语教学类研究因包含教师、教材、教法，类型较为丰富，所以方法上较其他类别文章相对多元。语料选取仍以 HSK 动态作文语料库为主，不同的是教学类文章通常会参考相关文件，比如词汇研究会参照《HSK 词汇等级大纲》，语法研究通常会参照《国际汉语教学通用课程大纲》等。在教学（法）的研究上主要集中于课堂模式和教学方法的探讨，比如综合课中汉字的教学模式教材研究中，研究最为常见的方法是对比，将同一主题下的多本教材或词典进行比较，分析优劣，并给出改进意见。目前唯一的1篇关于教师的研究是教育行业用语泛化研究，该文的研究方法是对行业用语的结构、语义特征进行描写，本质上仍属于使用三个平面理论对文本进行分析的方法范畴。

在文化与其他类的研究中，使用最多的是自建语料库，或是从已有语料库中选取一部分贴合研究主题的语料，这其中包含一些国际新闻报道中的语料，选取其中符合要求的内容，再借用相关文化理论进行分析，以探讨文化认同、国家形象、文化内涵及相应教学对策等话题。比如以文化认同理论为基础，运用话语分析和语料库分析相结合的方法，研究印尼华文媒体报道中的中国文化认同情况及话语建构策略（陈晓林，2021）。其语料的选取是来自对应国家新闻报刊的专栏，如印尼《国际日报》“中国新闻”专栏报道等。

表 5. 各类研究中语料库使用频次表

北京语言大学 HSK 动态作文语料库	15 次	暨南大学书面语语料库	5 次
CCL 语料库	5 次	BCC 语料库	4 次
全球中介语语料库	2 次	调查问卷	2 次
国家语委现代汉语语料库	1 次	汉籍全文检索系统	1 次

除上述语料的选取和各小类会涉及的方法外，通常会用到一些软件也可以顺带提及，分词软件最为常用的是 Corpus Word Parser，语料的处理通常会使用 WordSmith 或 Antconc 等语料库检索软件，如果涉及语料的转写通常会使用 FileToFolder 先将音频转写成文档文件，再使用语宝标注或是熊子瑜开发的 xRecorder 等进行转写，转写时的标注一般会遵循会话分析学派的转写系统进行，但目前口语语料转写成文字时不同论文所遵循的标准都不唯一，尚无一个统一的标准。

4 使用语料库法进行专业学位学位论文写作中存在的不足

Tognini-Bonelli (2001) 总结语料库语言学研究有两种范式：一是基于语料库——也就是利用语料库检验现有理论/修正，二是语料库驱动——也就是对各类现象进行全新界定。目前国际中文教育专硕生在撰写研究论文时，比较偏向于前者，也就是把语料库当成一种数据的来源。论文作者更倾向于利用语料来检验或修正现有理论，这是一种程式化的论文写作思路，比如在对一组同义词进行辨析搭配研究时，通常会就 MI 值、T 值进行检验，会围绕语义韵、节点词、类联接等方面展开论述，这本身无可厚非。基于对论文摘要及后记的分析，我们提倡国际中文教育专硕生在撰写毕业论文时，不仅仅要考察上述相关必要维度，在此基础上还要选取一些可用的理论，比如可懂输入、中介语理论等等，再次结合进行分析，然后针对不同的应用

场景给出一些教学建议/理论修正,这样的论文和研究才能真正对实际教学有所帮助。当然,这样的不足也体现为不同选题中的文章,语料的选取来源和使用的研究方法几乎相同。通常我们认为研究方法的选取是由研究目的来决定的,在五大类专硕学位论文中,各小类文章所使用的方法几乎一致。三个平面理论的使用固然有其价值,但是不同的研究目的还是有待做进一步区分,比如韵律语法的考察应该从韵律组配的角度出发,考察其韵律格式,语用的文章应该重点凸显语用上的差异,语法类的文章应该在句型模式、语法功能上做更深入地探讨。

此外就研究工具而言,定量分析通常需要作者在语料库中搜索一定数量的相关语料,并做数据的统计分析,也有一些选取了比较冷门的,比如 Sketch Engine 等。工具本身并无好坏,但是不同研究目的的文章,选取什么样的研究工具更加合适,需要对软件的功能有清楚的认知。目前国际中文教育领域专硕使用最多的语料库量化统计工具是 Wordsmith 和 Antconc,有学者(王菲 & 王天竹, 2021)曾经将 Wordsmith 和 Antconc 做过对比,我们再自行归纳成表格的形式加以说明。

表 6. Wordsmith 和 Antconc 对比表

功能	Wordsmith	Antconc
主要功能	1. 检索功能:以节点词为中心,按研究者需求以一定跨距内的语境共现情况进行排列,可以考察语义韵、类联接、语义倾向等; 2. 单词列表:筛选出需要的词汇及其意义; 3. 词频统计:统计词出现频率高低、并可以进行排序; 4. 主题词:通过制作词频表,对词汇进行对比	1. 检索功能:检索对象可以是词、短语、句子; 2. 索引词图:以图示的方式表明词在文本中的位置及亲疏程度; 3. 词频统计:统计词出现频率高低、并可以进行排序; 4. 主题词:将文本进行对比,列出差异的词汇
性能	算法好,执行速度快	处理汉语时需要额外指令和操作
界面 & 操作难度	界面复杂,不易操作;	界面简单,易于上手
其他	付费、商业软件	免费

从表 6 不难发现,在一些基本检索上,二者没有太多区别,但是在语法检索、跨文档比较等领域, Antconc 较 Wordsmith 略显欠缺,且缺乏翻译研究所需功能。但同时,不可否认的 Antconc 好处也有很多,比如在考察词汇组合情况,或是衡量由单词组成的词组能否构成搭配时, MI 值、T 值的计算是更为便捷的。在研究所需费用上, Antconc 也具有优势。研究时可以根据实际情况和研究需求来选择相应工具。如果涉及相关口语语料的转写,在使用 Antconc 时还需要进行一些预处理,因为软件本身是为印欧语系的语言设计的,汉语书写本身不会词之间不会分开,所以使用一些其他软件如 Corpus Word Parser 人工处理就显得十分重要,这有助于提升文本处理的准确度。

5 结语

目前国际中文教育专业硕士学位论文写作中,使用语料库法的人数越来越多。专硕生因其培养目标不同,故而在选题上更偏向实践,论文写作侧重与教学实践有关选题的特点也十分明显。但当前专硕生在选题上仍存在极不均衡的现象,部分选题无人问津,对语料库的利用局限于使用语料库检验现有理论。研究方法上受分析统计工具局限,存在多样化选题但单一化操作的缺点。语料库的发展为专硕生的专业学习和发展提供了便利的工具,在对专硕生的培养环节应加强对专硕生利用语料库进行研究的训练。

参考文献

- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. John Benjamins Publishing Company.
- 黄伟、李珺婷(2022)近二十年汉语二语偏误分析硕士学位论文的计量分析,《云南师范大学学报》(对外汉语教学与研究版), 2: 22-33。
- 王辉、施武佳(2022)汉语国际教育专业硕士学位论文高被引情况的计量分析,《云南师范大学学报》(对外汉语教学与研究版), 2: 1-10。
- 李丽(2022)汉语国际教育专业硕士文化类学位论文选题与写作:困境与出路,《天津师范大学学报》(社会科学版), 2: 24-30。
- 黄微(2022)汉语国际教育专业硕士文化类学位论文选题研究——以中国知网收录论文为例,江西科技师范大学硕士学位论文。
- 徐晶晶(2022)近十年甘肃高校汉语国际教育硕士学位论文选题研究,《兰州交通大学》, 2:90。
- 鄢芹芹(2021)“吃”类说法的体系性研究,华中科技大学硕士学位论文。
- 石芳联(2021)基于语料库的宋、元动态助词“了”使用频次差异显著性检验,湖北工业大学硕士学位论文。
- 陈如岳(2016)对外汉语中“动词+处所介词短语”的结构研究,新疆师范大学硕士学位论文。
- 鲁健骥(1992)偏误分析与对外汉语教学,《语言文字应用》, 1: 69-73。
- 周小兵(2017)汉语国际教育专业硕士毕业论文的研究设计与写作(上),《国际汉语教育(中英文)》, 1: 25-32。
- 陈晓林(2021)印尼华文媒体中国文化认同的话语建构策略——以印尼《国际日报》,中国石油大学硕士学位论文。
- 王菲、王天竹(2021)语料库检索软件的功能与特点对比研究——以 WordSmith Tools 6.0 与 AntConc3.2.4 为例,《信息与电脑》(理论版), 17:10-12。

投稿: 2024年1月2日; 接受: 2024年2月20日; 出版: 2024年4月7日

作者简介

王昌宇, 同济大学研究生, 南开大学国际教育学院教师, 研究方向: 国际中文教育。

刘运同, 同济大学教授, 研究方向: 国际中文教育, 语料库语言学。

An Investigation of Postgraduate Students' Thesis Writing Based upon the Corpus of International Chinese Education

Changyu Wang

Yuntong Liu

Tongji University, China

Abstract

Based on the search of academic platforms such as CNKI and DUXIU, this paper made a quantitative investigation on the theses of postgraduates majoring in international Chinese Education (including “Teaching of Chinese to Foreigners” and “Teaching of Chinese as a second language”) from 2010 to 2023, focusing on the topic selection and research methods of thesis writing. Using tableau, CiteSpace and other data analysis softwares and through an objective and subjective analysis, this paper analyzed the current research status of various topics, identified the problems and deficiencies that need our attention, and finally put forward corresponding suggestions for improvement.

Keywords

Corpus, professional masters' theses, selective topics

Changyu Wang is a postgraduate student at Tongji University and a teacher at the School of International Education, Nankai University, with research interests in international Chinese language education.

Yuntong Liu is a professor at Tongji University, with research interests in international Chinese language education and corpus linguistics.