

基于对比的中高级水平汉语二语学习者笔语词汇书面语特征的多维度考察

张江丽

北京华文学院, 中国

摘要

本文基于两大自建语料库——“外国留学生汉语笔语语料库”和“汉语母语者笔语语料库”，从完全书面语词汇、第一/第二人称代词、语气词这三个参数考察了中高级水平汉语二语学习者笔语词汇书面语特征的情况，并与汉语母语学习者进行了对比。研究表明：(1) 中高级水平汉语二语学习者笔语词汇的书面化程度远不及汉语母语学习者。(2) 高级水平汉语二语学习者笔语词汇的书面化程度高于中级水平学习者。(3) 与中级水平汉语二语学习者相比，高级水平汉语二语学习者在完全书面语词汇的使用数量上增长较为显著；而高级水平汉语二语学习者在第一/第二人称代词和语气词的使用数量上均略有下降。

关键词

笔语词汇, 书面语特征, 语料库, 汉语母语者, 汉语二语学习者

1 引言

在汉语第二语言的教学实践中，我们发现很多学生“会说不能写，能聊看不懂”。其中一个非常重要的原因就是汉语二语学习者缺乏语体意识，对汉语口语和书面语之间的差异不甚了解。

汉语的语体大致可分为正式的书面语体和非正式的口语语体。不同语体的特征又体现在语音、词汇、语法、修辞、篇章等诸多方面。本研究将基于大规模语料库从词汇的角度考察汉语二语学习者笔语词汇书面语特征的习得情况。

2 概念界定及相关文献回顾

2.1 相关概念界定

2.1.1 “笔语”与“口语”

“笔语”与“口语”是一组相对的概念。“笔语”是笔头语言，是指用纸笔书写的形式在书面记录下来的语言材料，如作文、日记、书信、留言等。“口语”是口头语言，是指用口头表达的方式输出的语言材料，如访谈、录音等。“笔语”与“口语”是从语料来源的角度对语言材

料进行的分类。需要说明的是,一些访谈、录音虽然后来被用纸笔的形式转写记录下来,但究其根本,其语料性质仍为口语语料。本研究中所使用的“笔语”语料均为学生以纸笔书写的方式写成的作文。

2.1.2 “书面语”

语体是“一种言语体式”(黄伯荣、李炜,2016),是“由语言运用特点所形成的体系,是民族共同语的功能变体”(袁晖、李熙宗,2005)。在现代汉语有关语体的研究中,张弓的《现代汉语修辞学》最早把语体分为口头语体和书面语体两大类。之后多本比较有影响力的教材,如:胡裕树主编的《现代汉语》、张静主编的《新编现代汉语》等大都沿用了“二分法”这一分类方法。也有学者采用了“三分”的分类系统(冯胜利,2018),在书面语体和口语语体中间加入了中间语体。

本研究所说的“正式的书面语体”就是指“二分法”中的“书面语体”,即在相对正式的场合中使用的、以书面表达为典型形式(主要为书面语词和书面句式)的一种语体。

2.1.3 “现代汉语书面语”

现代汉语书面语是在相对正式的语言环境中使用的、以书面表达为典型形式的一种语体,现代汉语书面语具有典雅、规范、庄重的风格特点。跟口头语体相比,作者在生成书面语体的过程中,通常有更多的思考时间,因此,在词语方面,正式的书面语体往往选择典雅、庄重的词语;在句子方面,书面语体也更加完整、规范。

2.1.4 笔语词汇的书面语特征

“笔语词汇的书面语特征”是本研究的核心问题。简言之,本研究的目标是考察汉语二语学习者在以纸笔书写的方式写成的作文中书面语体词汇特征的表现情况,也可称之为“书面化程度”。

2.2 相关文献回顾

2.2.1 二语词汇语体特征衡量参数的相关研究

在英语第二语言教学领域,学者们对词汇语体特征的考察集中在作者/读者显现度(Petch-Tyson, 1998; 文秋芳等, 2003)、不同词频等级的词汇分布(Cobb, 2003; 文秋芳等, 2003)、第一/第二人称代词使用的多寡(文秋芳, 2009; 潘璠, 2012)等方面。研究普遍表明我国英语学习者作文中存在明显的“口语化”倾向。

在汉语语体研究领域,冯胜利等(2008)提出把嵌偶词、合偶词、书面句式、《汉语水平词汇与汉字等级大纲》中的丁级词作为衡量词汇语体特征的参数,研制出了测量庄雅度的公式。他首次从量化的角度探讨了汉语二语学习者词汇的语体问题,具有开创性的意义。郭锐(2002)研究发现现代汉语语气助词、叹词、拟声词具有明显的口语语体色彩。宋婧婧(2013)基于大规模的有声媒体和平面媒体语料库,对口语和书面语词汇进行了定量研究,研究发现口语中更多地使用了代词和语气词。马明艳(2017)以语气词、第一/第二人称代词、介词为参数采用语料库数据对比分析的方式,考察了汉语作为第二语言学习者作文的词汇使用状况。冯胜利等

(2021)研究表明汉语中的单双音节对应词在语体上具有对立属性。王用源、陈宇豪(2023)从语体语法的视角提取了10项评价语体正式度的特征指标,构建了语体正式度的测量模型,其中与词汇语体相关的指标有词长、嵌偶词词频、合偶词词频、词汇复杂性、动态助词词频这5个指标。

综合以上与词汇语体特征有关的研究,我们可以发现,衡量词汇语体特征的参数大致可以分为两类:(1)直接参数。直接参数是指具有书面语特征的词汇。直接参数是衡量词汇语体特征最直观、最有效的参数。冯胜利等(2008)提出的参数中的嵌偶词、合偶词、书面语句式就是直接参数。(2)间接参数。间接参数是衡量词汇语体特征的补充性参数。通常是某些特殊的词类或词汇的频度分布。如:语气词、第一/第二人称代词、介词、叹词、拟声词、动态助词、词长、不同词频等级的词汇分布等。

在已有的研究中,仅有冯胜利等(2008)的研究涉及到了直接参数。这一研究引发了我们的思考,现代汉语中具有书面语特征的词汇不仅限于冯胜利等(2008)所提出这些特殊词语和句式。如果能把这部分词汇相对穷尽,归纳出具有典型书面语特征的现代汉语词汇,将在一定程度上推进现代汉语词汇书面语特征的研究。现代汉语有关书面语词汇的专门工具书以及一些权威工具书中标注有<书>的词语也是具有典型书面语特征的词汇。本研究拟在搜集这些资料的基础上,归纳总结出具有典型书面语特征的词汇,我们称之为“完全书面语词汇”,将其作为本研究衡量词汇书面语特征的重要参数。“完全书面语词汇”的选取过程见下文。

在已有关于词汇语体特征间接参数的研究中,以汉语二语学习者为考察对象的仅有马明艳(2017)的研究。这项研究选取语气词、第一/第二人称代词、介词为参数进行考察。由于介词是汉语二语学习者学习的易错点,因此在我们搜集到的汉语二语学习者的语料中,学习者出现的介词偏误相对较多。为了避免学习者偏误对词汇语体特征的考察产生影响,我们舍弃了介词这一参数。在间接参数中,本研究拟选取第一/第二人称代词、语气词来考察第二语言学习者词汇的语体特征。

在总结前人研究的基础上,我们确定了衡量笔语词汇书面语特征的三个参数:(1)完全书面语词汇;(2)第一/第二人称代词;(3)语气词。

2.2.2 词汇语体特征测量方法的相关研究

在冯胜利等(2008)研制出的衡量语体庄雅度的公式中,我们发现作者直接把嵌偶词作为测量语体庄雅度的参数,这样稍有不妥。因为从冯胜利(2006)在《汉语书面语初编》一书中对嵌偶词的定义来看,嵌偶词是单字词,作者在书中举例:“暗”“备”“错”“力”为嵌偶词,嵌偶词仅在组成特定的双音词时才具有较强的书面语特征,例如:“错”在组成“错杀”“错认”“错砍”时才具有较强的书面语特征,而在组成“错误”时并无明显的书面语特征。所以直接把“嵌偶词比例=嵌偶词出现次数/字数”作为测量语体庄雅度的公式之一,是存在问题的。因为如果只是把“错”这样的单字词作为衡量标准的话,很多类似“错误”这样的虽然包含嵌偶单字“错”字,但是却没有明显的书面语特征的嵌偶词会被计算在内,这样并不能反映语体的正式度。因此,分子改为“嵌偶词所组成的书面语词汇的出现次数”更为合适。与之对应的分母,应改为与分子单位一致的词汇单位——“词形数(即不重复出现的词的数量)”。因此,该公式改为“嵌偶词比例=嵌偶词所组成的书面语词汇的出现次数/词形数(即不重复出现的词的数量)”较为合适。

还有学者对比了不同语料库中语料的语体特征,调查了某些词类在几种语料中的差异。如马明艳(2017)以“HSK动态作文语料库”“国家语委现代汉语语料库”“北京口语语料库”为语料来源,通过检索的方式对比了语气词、人称代词、介词在这三个语料库中分布,进而得出汉语二语学习者的作文存在“口语化”倾向的结论。在这一研究中,作者选取的“北京口语语料库”

主要是采用谈话和录音的方式收集了 500 个土生土长的北京人的口语语料。作者将这一语料与另外两个语料库中的语料进行了对比。我们认为这一语料库与其他两个语料库的语料性质差别较大,不适合直接对比。

在汉语中,用文本形式呈现的语料通常有两种:一是采集的语料本身就是用纸笔的形式书写的,这类语料后来被以文本的形式录入电脑,这是典型的笔语语料。二是采集的语料本身是以录音的形式存储的,后期通过软件或人工把录音转写成了文字录入电脑。这类语料虽然从最终的呈现形式上与第一种语料相同,但从其来源来看,其本质是仍然是口语语料,这类语料我们可以称之为“笔语化的口语语料”。马明艳(2017)选取的“北京口语语料库”就属于这种类型。因此,马明艳(2017)所选取的“北京口语语料库”与其他两个语料库的语料不是同一性质的语料,可比性不强。

综上所述,以往面向汉语二语学习者的词汇语体特征的研究相对较少。研究所涉及的笔语词汇书面语特征的参数较为单一,要么只选取直接参数,要么只选取间接参数,尚未看到将直接参数和间接参数结合起来综合衡量笔语词汇书面语特征的多维度研究。在衡量书面语词汇的直接参数的选取上,已有研究仅限于嵌偶词、合偶词、书面语句式等相对零散的参数,缺乏对已有书面语词汇大规模、穷尽式的筛选和综合。在已有采用间接参数开展的研究中,在进行对比研究时,选取的对比语料库中的语料规模、语料性质不一致,缺乏同质性。因此,基于大规模语料库,多维度考察汉语二语学习者笔语词汇书面语特征的研究很有必要开展。

2.3 研究目标

本研究拟以两大自建语料库——“外国留学生汉语笔语语料库”和“汉语母语者笔语语料库”为语料来源,运用对比的方法,考察笔语词汇书面语特征的各个参数在两大语料库中的差异,进而考察中高级水平汉语二语学习者笔语词汇书面语特征的使用情况及其与汉语母语学习者的差异。本研究主要聚焦于以下两大问题:

第一,中高级水平汉语二语学习者与汉语母语学习者在笔语词汇书面语特征的三个参数——完全书面语词汇、第一/第二人称代词、语气词上的分布有何差异?

第二,中高级水平汉语二语学习者笔语词汇书面语特征的三个参数——完全书面语词汇、第一/第二人称代词、语气词的发展趋势如何?

3 “完全书面语词汇”的选取及相关问题说明

3.1 现代汉语“完全书面语词汇”的选取

无论在汉语中还是英语中,相邻语体之间的界限并非泾渭分明,但是典型的书面语体和典型的口语语体之间的差别是显著的。Biber(1988)认为,口语语体与书面语体形成了一个连续统,其他的语体都处在从口语语体到书面语体的连续统上。虽然连续统是渐变的,相邻的语体差异是模糊的,但是位于连续统两端的语体差异显著,界限清晰。冯胜利(2010)认为“语体必两极对立而后存”。曹炜(2003)就把书面语词分为两大类:一类是完全书面语词,一类是不完全书面语词。完全书面语词相对封闭,有据可查,具有可操作性,因此,完全书面语词可以作为衡量学习者笔语词汇书面语特征的重要参数。现代汉语中有关书面语词汇的工具书和词典中

标有<书>的词汇是确定完全书面语词汇的重要依据，它们也让笔语词汇书面语特征的计量分析成为可能。

我们查找了各大现代汉语书面语词表，对工具书中的书面语词汇进行了汇总。目前查找到的工具书和词表如下：

表 1. 书面语词汇工具书

工具书	编著者	出版社	出版时间	词条范围	不重复的词条数量（个）
《中华书面语词典》	王安节、鲍海涛	吉林文史出版社	2003	所有词条	3305
《汉语书面用语初编》	冯胜利	北京语言大学出版社	2006	嵌偶词组成的双音词，合偶词	1153
《现代汉语词典》第六版	中国社会科学院语言研究所词典编辑室	商务印书馆	2012	标注<书>的所有词条	5559

我们首先将以上工具书中所列书面语词汇录入 excel 表格。对于以下几种情况，在录入时进行了特殊处理：

（1）词的前后有一定使用条件的词条。例如：毕，只有在动词后面单用，才是书面语；对于这种情况，我们在录入时，将其录入为“V+毕”。

（2）一些词有多个词性，仅用作某种词性时方为书面语词。如：“及”只有做连词时才是书面语词，而由“及”组成的词——“及时”“来得及”等并非书面语词。再如：“左右”在用作动词时才是书面语词。由于我们的语料都经过了词性标注，因此检索时，软件只识别其用作书面语词的词性。

经录入、筛选、去重后，我们发现工具书和词表中涉及的书面语词汇包括以下三种类型：

一是完全书面语词或短语，如：聆听、绝伦、劳顿。该类型占绝大多数。

二是书面语组合，例如：备+V；避+N；遍+V。

三是书面语关联搭配。例如：为……所困；以……为乐。

统计显示以上工具书中所列不重复的现代汉语书面语词汇共计 9046 个，其中完全书面语词 8980 个，词语组合 63 个，关联搭配 3 个。这些词汇形成了《现代汉语完全书面语词汇总表》。

综合前文论述，本研究最终确定了笔语词汇书面语特征的测量参数、测量单位及计算方法。可用下表表示：

表 2. 汉语笔语词汇书面语特征测量参数及计算方法

测量参数	测量单位	计算方法
参数 1： 完全书面语词汇	《现代汉语完全书面语词汇总表》中的完全书面语词、书面语组合、关联搭配的产出数量	百万语料中出现的次数
参数 2： 第一/第二人称代词	第一/第二人称代词的产出数量	百万语料中出现的次数
参数 3： 语气词	语气词的产出数量	百万语料中出现的次数

在以上三类参数中,参数1为正向特征,参数2、参数3均为负向特征。我们的基本假设是:参数1“完全书面语词汇”在语料中出现的数量越多,比例越大,学习者笔语词汇的书面语特征就越显著。参数2“第一/第二人称代词”和参数3“语气词”在语料中出现的数量越多,学习者笔语词汇的书面语特征就越不显著。

3.2 词汇语体特征分析软件的设计及分词软件的选取

本研究请专业的计算机工程师设计了面向笔语词汇书面语特征研究的词汇分析软件。对于《现代汉语完全书面语词汇总表》中的第一种类型——“完全书面语词”实现直接检索覆盖;对于第二种类型——“书面语组合”和第三种类型——“书面语关联搭配”,通过词性标注,设置检索条件,实现检索和覆盖。

在软件设计过程中,分词问题是影响词汇统计结果的关键要素。我们对多家较有影响的分词软件进行了试运行,综合考虑分词数量和分词正确率等因素,最终选用中国传媒大学国家语言资源监测与研究有声媒体中心开发的分词软件作为本研究的分词工具,在此基础上对分词结果进行了人工校对。

3.3 语料来源

本研究使用的语料包括汉语二语学习者的笔语语料和汉语母语学习者的笔语语料。

本研究使用的汉语二语学习者的语料来自于自建语料库——“外国留学生汉语笔语语料库”。该库所采集的语料均为在中国学习汉语的外国留学生平时的习作以及期中、期末考试状态下的作文。为了更真实地反映学习者的汉语水平,所采集的语料要求必须是学习者独立完成的、且未经他人指导和修改的原始语料。截止目前,语料库规模已达到4628篇,约188万字。目前语料库中的学习者来自印尼、泰国、美国、日本、德国、英国、马来西亚、韩国、菲律宾、蒙古、老挝、柬埔寨、俄罗斯等42个国家。学习者的汉语水平以学时为标准进行划分,初级水平学习者的学时少于960小时,中级水平学习者的学时在960~1920小时之间,高级水平学习者的学时在1920小时以上。我们从自建的“外国留学生汉语笔语语料库”中选取了16岁以上的中高级水平学习者的语料100万字,共计2146篇。

本研究使用的汉语母语学习者语料来自于自建语料库——“汉语母语者笔语语料库”。该语料库中的语料均为母语为汉语的中国学生所写的课堂作文及考试作文的原始语料。截至目前,语料规模已达200万字。我们从中选取了高中三年级学生的语料100万字,共1600篇。为避免文章体裁对两类学习者词汇使用的影响,我们从两大语料库中选取的文章均为记叙文。

本研究之所以选择中高级水平汉语二语学习者和高中三年级学生的作文主要有以下两个方面的考虑:一是中高级水平汉语二语学习者的作文和高中三年级学生的作文中的语言已经基本成熟。二是这两大语料库中语料产出者的年龄相近,思想深度与思维能力相当,在写作内容上具有一定的可比性。虽然这两类学习者在汉语写作水平上有一定的差距,但是本研究关注的焦点是,中高级水平汉语二语学习者与汉语母语学习者在笔语词汇书面语特征上的异同。从这两类语料大致可以看出两类学习者的差异,能基本满足研究的需要。

4 汉语二语学习与汉语母语学习者笔语词汇书面语特征情况对比

4.1 完全书面语词汇在两大语料库中的分布情况对比

完全书面语词汇包括书面语词、书面语组合、书面语关联搭配这三种情况。我们分别对此进行了考察。

4.1.1 书面语词的使用情况对比

汉语二语学习和汉语母语学习者书面语词的使用状况如下表：

表 3. 汉语二语学习与汉语母语学习者书面语词覆盖情况

学习者	书面语词 使用数量（个）	《书面语词汇总表》中书面 语词总数（个）	使用比例 （%）
汉语二语学习者	1554	8980	17.3
汉语母语学习者	3008	8980	33.5

在 100 万字的汉语二语学习者的笔语语料中，学习者使用的词形数（即不重复出现的词的数量）共有 20913 个，其中汉语书面语词共使用了 1554 个，书面语词的使用数量占全部词形数的 7.4%，占《书面语词汇总表》的 17.3%。

在汉语母语学习者 100 万字的语料中，母语学习者使用的词形数共有 29107 个，其中汉语书面语词使用了 3008 个，书面语词的使用数量占全部词形数的 10.3%，占《书面语词汇总表》的 33.5%。通过表 3 的数据对比可以发现，汉语二语学习和汉语母语者在书面语词的使用数量上有较大差距，相差将近一倍。可见，在本研究考察的两类笔语语料中，汉语母语学习者书面语词的使用比例相对较高，二者之间的差距较为明显。

4.1.2 书面语组合的使用情况对比

同样，我们对归纳出来的《完全书面语总表》中的第二类——书面语组合（如：备 +V；避 +N；遍 +V）的使用情况进行了统计分析，结果如下表：

表 4. 汉语二语学习与汉语母语学习者书面语组合覆盖情况

学习者	书面语组合使用数量（个）	《书面语词汇总表》中词语 组合总数（个）	使用比例（%）
汉语二语学习者	28	63	44.4
汉语母语学习者	35	63	55.6

从上表可以看出，汉语二语学习者对书面语组合的使用状况与汉语母语学习者有一定差距，汉语二语学习者使用的词语组合的数量约为汉语母语学习者的 80%。

对于《现代汉语书面语词汇总表》中的第三类——书面语关联搭配，在这两类语料库中均未出现。在此不做讨论。

4.2 第一 / 第二人称代词在两大语料库中的分布情况对比

我们对两大语料库中第一 / 第二人称代词的出现情况进行了统计，结果如下：

表 5. 汉语二语学习与汉语母语学习者第一 / 第二人称代词的出现次数对比

学习者	我	我们	你	你们	合计
汉语二语学习者	29846	4618	4111	736	39311
汉语母语学习者	11681	3325	4304	150	19460

统计结果可以看出，汉语二语学习者第一 / 第二人称代词的出现次数明显多于汉语母语学习者，汉语二语学习者第一 / 第二人称代词的使用总数约为汉语母语者的 2 倍。按照使用频次，汉语二语学习者使用的人称代词从多到少依次是：我 > 我们 > 你 > 你们；而汉语母语学习者使用的人称代词从多到少依次是：我 > 你 > 我们 > 你们。可见，汉语二语学习者优先使用第一人称代词，汉语母语学习者优先使用单数人称代词。

4.3 语气词在两大语料库中的分布情况对比

表 6. 汉语二语学习与汉语母语学习者语气词的出现次数对比

学习者	吗	啊	吧	呢	合计
汉语二语学习者	639	357	545	617	2158
汉语母语学习者	672	628	593	584	2477

研究表明：汉语母语学习者的语料中语气词的使用数量略多于汉语二语学习者的使用数量。

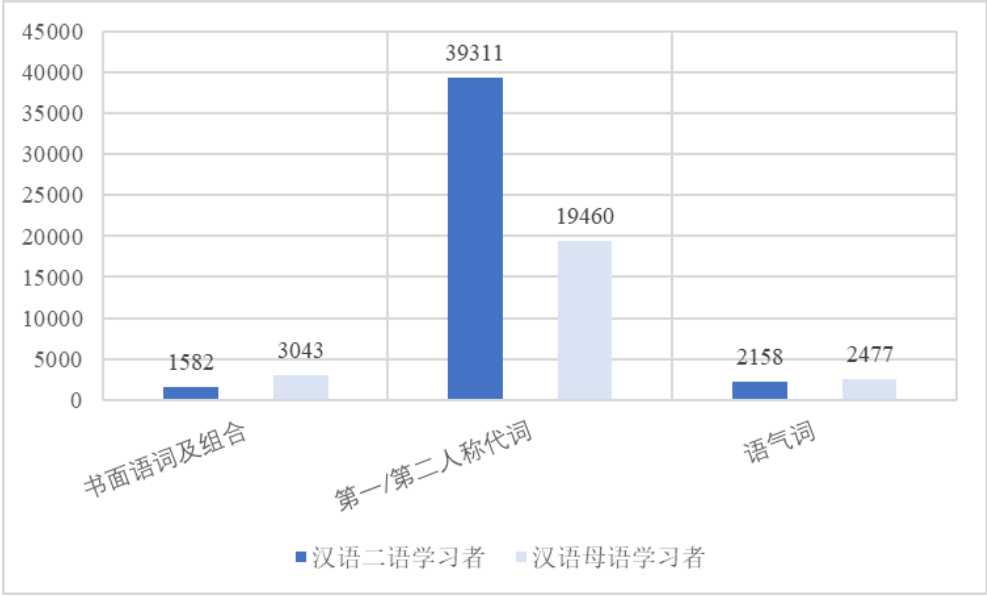
郭锐（2002）研究发现语气助词、叹词、拟声词具有明显的口语语体色彩。宋婧婧（2013）基于大规模的现代汉语有声媒体和平面媒体语料库，对口语和书面语词汇进行了定量研究，研究发现口语中使用了更多代词、语气词。这两项研究均表明，在汉语母语者的语料中，代词和语气词具有较强的口语语体特征。马明艳（2017）对汉语二语学习者书面作文的词汇使用状况的研究表明，语体越正式，语气词的使用数量越少。

为什么我们的研究结果与以上研究不同呢？出现这一现象可能是因为汉语母语学习者在书面作文中，尤其是本研究所选取的记叙文的写作中，存在口语化的倾向，因此，汉语母语学习者的语气词使用相对较多。

4.4 汉语二语学习与汉语母语学习者笔语词汇书面语特征对比

汉语二语学习与汉语母语学习者在笔语词汇的书面语特征上有何异同呢？这两类学习者在笔语词汇的书面语特征三个参数上的表现如下图所示：

图 1. 汉语二语学习与汉语母语学习者书面语特征对比图



从两类学习者在笔语词汇的书面语特征上表现来看，在三个参数中，从柱形高度来看，汉语二语学习者在直接参数——书面语词及组合上的使用数量约为汉语母语学习者的二倍。这说明汉语二语学习者对书面语词汇的习得和使用状况远不及汉语母语者，汉语二语学习者的书面语特征不够显著。在间接参数上，汉语二语学习者对第一/第二人称代词的使用数量远高于汉语母语者，而在语气词的使用上，汉语二语学习者略低于汉语母语者。

5 不同水平汉语二语学习者笔语词汇书面语体特征的使用状况

5.1 不同水平汉语二语学习者完全书面语词汇的总体使用状况

完全书面语词汇包括书面语词、书面语组合、书面语关联搭配这三种情况。由于在语料中只出现了前两种类型，因此，我们分别对中级、高级水平汉语二语学习者书面语词、书面语组合的使用状况进行了考察。

5.1.1 书面语词的使用情况对比

在 100 万字的语料中，中高级水平汉语二语学习者使用的书面语词的情况如下：

表 7. 中高级水平汉语二语学习者书面语词的总体使用情况

汉语水平	出现次数	书面语词总数（个）	总表中的使用比例（%）
中级	755	8980	8.4
高级	1260	8980	14.0

从上表可以看出，与中级水平汉语二语学习者相比，高级水平汉语二语学习者使用的书面语词数量有较大增长。高级水平学习者使用的书面语词约为中级水平学习者的 1.7 倍。可见，随着汉语水平的提高，汉语二语学习者对汉语书面语词的掌握情况越来越好。

5.1.2 中高级水平汉语二语学习者书面语组合的使用状况

在 100 万字的语料中，中高级水平汉语二语学习者使用的书面语组合的情况如下：

表 8. 中高级水平汉语二语学习者书面语组合的使用情况

汉语水平	出现次数	书面语组合总数（个）	总表中的使用比例（%）
中级	19	63	30.2
高级	23	63	36.5

同样，中高级水平汉语二语学习者书面语组合的使用状况与汉语书面语词的使用状况相似，随着汉语水平的提高，学习者使用的书面语组合的数量也越来越多。不过由于汉语书面语组合的总数较少，因此，书面语组合的数量增长幅度不大。

5.2 中高级水平汉语二语学习者第一 / 第二人称代词的使用情况

表 9. 中高级水平汉语二语学习者第一 / 第二人称代词的使用情况

汉语水平	我	我们	你	你们	合计
中级	7285	1753	2328	288	11654
高级	7635	1596	1057	90	10378

从不同水平汉语二语学习者第一 / 第二人称代词的使用总数来看，中级水平学习者使用的数量高于高级水平学习者。

在这四个人称代词中，按照使用频次从高到低，中级水平学习者使用的代词数量依次为：我 > 你 > 我们 > 你们。而高级水平学习者使用的代词依次是：我 > 我们 > 你 > 你们。

5.3 中高级水平汉语二语学习者语气词的使用情况

表 10. 中高级水平汉语二语学习者语气词的使用情况

汉语水平	吗	吧	呢	啊	合计
中级	188	218	289	97	792
高级	222	179	198	179	778

从中高级水平汉语二语学习者语气词的使用情况来看，中级水平学习者使用的语气词数量略高于高级水平学习者。随着汉语水平的提高，汉语二语学习者的语气词使用数量出现了小幅下降。

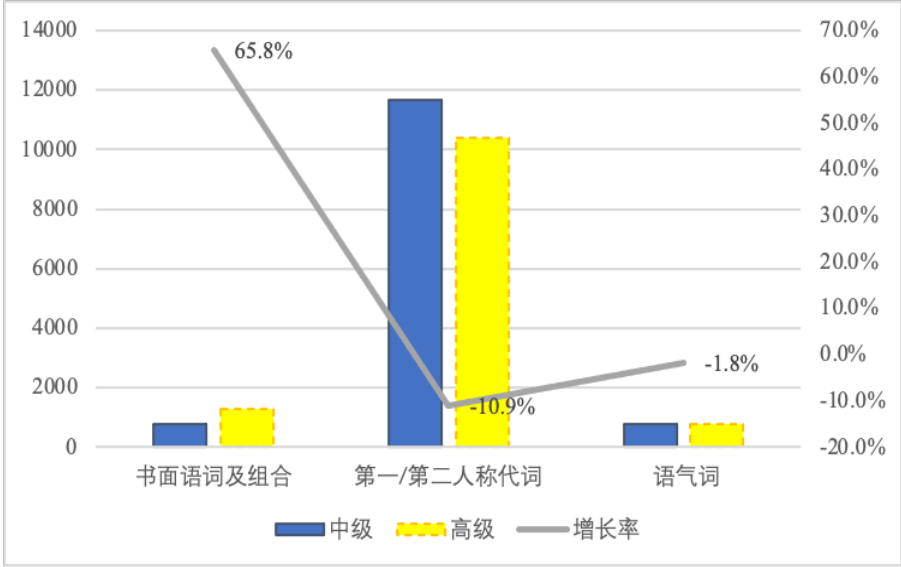
5.4 中高级水平汉语二语学习者词汇书面语特征的变化趋势

表 11. 各参数在中高级水平学习者语料中的出现次数

汉语水平	书面语词及组合	第一 / 第二人称代词	语气词
中级	774	11654	792
高级	1283	10378	778

随着第二语言学习者汉语水平的提高,在三大参数中,高级水平学习者使用的书面语词及组合上增加了 509 个,第一/第二人称代词减少了 1276 个,语气词减少了 14 个。由于语料中出现的 第一/第二人称代词的数量均在万次以上,所以从中级水平到高级水平,汉语二语学习者使用的 第一/第二人称代词的数量虽然减少了 1276 个,但是下降的幅度并不大。

为了更好地展示中高级水平汉语二语学习者在这些参数上的发展趋势,我们用折线图表示如下:



从上图可以看出,在本研究涉及的三个书面语特征的参数中,完全书面语词汇(书面语词及组合)、第一/第二人称代词、语气词随着汉语水平的提高,变化趋势不同。随着学习者的汉语水平从中级迈进高级,学习者使用的完全书面语词汇大幅增长,增长幅度达到了 65.8%。可见,中高级阶段是书面语词汇习得的重要时期。与此同时,第一/第二人称代词和语气词的使用数量均呈现下降趋势。第一/第二人称代词的使用数量下降了 10.9%,语气词的使用数量下降了 1.8%。

整体来看,高级水平学习者的书面化程度高于中级水平学习者。

5.5 中高级水平汉语二语学习者汉语书面语词汇出现频次分布

语料库中学习者使用的汉语书面语词汇的频率也不一样,有些书面语词在语料库中出现得较为频繁,有些书面语词汇却出现得较少。我们将语料库中书面语词汇(书面语词及组合)出现的次数从高到低分为四个区间:100 次以上、10 ~ 99 次、2 ~ 9 次、1 次。在此基础上,对中高级水平学习者汉语书面语词汇(书面语词及组合)的出现频次进行了分段统计,统计结果见下表。

表 12. 中高级水平学习者汉语书面语词汇出现频次分布

汉语水平	100 次以上	10 ~ 99 次	2 ~ 9 次	1 次	合计
中级	5	106	310	334	755
高级	3	157	578	522	1260
增长率	-40.0%	48.1%	86.5%	56.3%	66.9%

从上表可以看出,中级、高级水平学习者使用的书面语词汇总数分别为755个、1260个。从他们使用的书面语词汇的总数来看,高级水平学习者占绝对优势。在出现次数从高到低的四个区间中,中级水平学习者产出的书面语词汇集中在出现频次为1次的这个区间,有334个词汇仅出现了一次。可见,中级水平学习者使用的书面语词汇中低频次的词汇较多,书面语词汇的产出整体上较为分散。而高级水平学习者的语料中,学习者产出最多的是2~9次的书面语词汇,这样的书面语词汇有578个。可见高级水平学习者产出的书面语词汇相对比较集中,一些书面语词汇开始不断重复出现。在“2~9次”和“1次”这两个频段中,高级水平学习者的增幅较大,这说明高级水平学习者习得汉语书面词汇的数量也越来越多,这从某种程度上也证明了他们的汉语水平较高。

总体来看,与中级水平学习者相比,高级水平学习者对书面语词汇的熟悉度在不断提升,一些书面语词汇的使用频次明显增加,产出情况较好。

6 研究启示与建议

6.1 重视语体在第二语言教学中的地位和作用

第二语言教学的根本目的是培养学生的语言交际能力。李泉(2003)认为培养学生的语言交际能力,其本质上说就是培养学生运用目的语语体的能力。第二语言学习者能够在不同的场合和语境中选择合适的语言进行交际,离不开语体能力的培养。在汉语第二语言教学中,学生能否正确使用语体,是学生汉语水平高低的一个重要表现。在某种程度上,语体是区分学生汉语水平的重要标志。如果一个学生能在不同的场合正确得当地使用汉语的不同语体,那么这个学生的汉语水平通常较高。因此,语体教学和语体研究在汉语二语教学中占有十分重要的地位。

在汉语本体研究中,有关语体的研究成果较为丰富。但是本体理论研究与汉语第二语言教学之间还有一定的距离。很多本体有关语体研究的成果并未应用于汉语第二语言教学中。今后,应重视汉语本体中有关语体的研究成果向汉语二语教学的转化。

在汉语第二语言教学中,我们应该对“现代汉语书面语”进行明确的界定。现代汉语书面语在词语方面,往往选择典雅、庄重的词语;在句子方面,也更加完整、规范。

在汉语第二语言学习的过程中,学习者通常有这样的误区:他们以为只要把自己说的语言记录下来,就形成了现代汉语书面语。在教学中,教师要提醒学生区分什么是口语语言的笔语化,什么是真正的书面语。

在平时学生的习作中,我们也注意到很多初中级水平甚至高级水平的留学生作文中普遍存在的“我手写我口”的情况。这种作文基本上以口语化的词汇为主,语体意识淡薄。对于初级水平的学习者,我们鼓励学生大胆写,勇于尝试使用学过的各种词汇。在写作指导中,从最常见的词语搭配的角度入手,提醒学生注意不同词汇的搭配习惯和搭配条件,让学生体会到意思相近或相关的词语在不同的组合和搭配中的不同语体功能,引导学生具有初步的语体意识。在中级阶段的汉语第二语言教学中,我们应该从词汇、语法等方面将具有书面语特征的语言要素进行归纳整理和总结,让学生在写作和生活中能够正确区分和使用。经过前两个阶段的铺垫,到了高级阶段,学生应该有明确的语体意识,并有较强的语体应用能力,能在不同场合的语言表达和不同类型的文章中正确使用合适的语体。

6.2 重视词汇的语体特征, 分类制定语体词表

本研究表明:与汉语母语学习者相比,中高级水平汉语二语学习者笔语词汇书面语特征参数的产出情况并不理想,中高级水平汉语二语学习者笔语词汇的书面化程度远不及汉语母语学习者。

这一结果与汉语二语教学界对词汇语体特征的关注程度密切相关。在汉语二语教学中,有各种各样的大纲。李泉(2003)认为“现有的各类语法大纲总的说来都缺乏语体意识。具体说来,无论是词汇语汇还是句法格式句式(包括单句句式和复句句式)等都没有显示语体属性”。已有的各类大纲以“通用性”的语法和“中性语体”的词汇的教学为主。在汉语二语学习者学习汉语的初级阶段,首先要习得的应该是中性语体的词汇。但是到了中高级阶段,仍然只停留在中性语体范畴,不注意词汇的语体特征,很难培养出可以应付各种场合且语用得体的学生。正如S·皮特·科德(1983)所言:“如果一个学生只懂得一种代码的所有用途共有的东西,那么他在任何场合都不会恰当地使用这种代码”。目前汉语二语教学界缺乏从语体特征的角度入手,对汉语二语教学中的常用词汇,尤其是具有明显书面语或口语特色的词汇进行语体特征分类的词表并未出现。

本研究综合了本体研究中的书面语词汇表,研制出了《现代汉语完全书面语词汇总表》,我们可采用同样的方法对现代汉语中的完全口语词汇进行归纳汇总。在此基础上对完全书面语词汇和完全口语词汇进行分级,在编写教材时,按照学习者的汉语水平将不同等级的书面语词汇和口语词汇加入教材。日本仓石武四郎在《中国语辞典》(1963)中把双语词汇分为11级。一般语词为0级。从“文”到“古”分为上五级,从“俗”到“土”分为下五级。汉语第二语言教学大纲在对书面语词汇和口语词汇进行分级时,可参考仓石武四郎的分级方法进行调整。

6.3 加强口语词和书面语词的转换练习

本文的研究表明,中高级水平学习者在记叙文的写作中存在着明显的“口语化”倾向。这与二语学习者不了解汉语近义词或近义语素的文白对应关系有关。王永娜(2012)的研究表明双音化不仅是“形态的标记”,同时也是“语体转换的手段”。在学生日常的作业中,我们也经常发现一些“文白”混用的情况。如:“进行改”“举行会”“得成绩”。这种情况下,教师可引导学生进行“组单为双”的练习,把“改”替换为“修改”,“会”替换为“会议”,“得”替换为“取得”,逐步让学生感受到汉语词汇的音节特点与词汇语体的对应关系。

目前的汉语综合课、写作课、阅读课中均涉及到应用文(如:通知、告示、请假条、邀请函、演讲稿)的教学。大部分教材一般只列出了基本格式,对应用文中具有明显书面语体特征的词汇关注不够。应用文是汉语中比较典型的书面语体。这部分内容看似篇幅短小,但对学习者来说,难度并不小。例如:张世涛,刘若云编著的《初级汉语阅读教程I》中出现过一篇告示,内容如下:

告示

因线路维修,28日(星期三)上午8点到下午4点停止电力供应,请大家互相告知。特此通知。

水电科

2003年1月30日

这篇告示突出的难点就在于如何理解其中出现的一些书面语词汇。在课堂上,教师可以通过口语词和书面语词的转换,化解难点。如在讲解这条告示的时候,可以引导学生把偏书面语

色彩的词语“因”“维修”“停止”替换成偏口语色彩的词语“因为”“修”“停”。通过这种“变文为白”的操练,帮助学生理解告示的内容。为了强化应用文中书面语词汇的使用,教师还可以引导学生进行反向练习。教师可以用口语词汇来叙述这条告示的基本内容,然后让学生改写成较为正式的告示。通过“文”“白”的双向转换练习,帮助学生理解和运用应用文中的书面语词汇,同时让学生了解增减音节是改变词汇语体色彩的一种方法。

6.4 重视教材中词汇语体特征的标注与讲解

一般的汉语二语教材通常附有生词表。生词表一般只列出了拼音、汉语释义、英文释义、词汇等级等。对于词汇的语体特征,鲜见标注。目前可及的教材中,仅有李晓琪主编《博雅汉语》和冯胜利主编的《汉语综合教程》对典型的书面语词汇进行了标注。到了中高级阶段,词义辨析是词汇教学的重要内容。在近义词辨析中,词汇的语体特征是区分近义词的一个重要方面,学习者常常因为分辨不清词汇的语体特征而出现误用词语的情况。建议在生词后加入词汇语体特征的标注,这样可以有效地帮助学生提高正确使用词汇的能力。

在释义时,经常出现词汇意义对应但语体不对应的情况。例如,“亦”被翻译为“also”,“也”也被翻译为“also”。这种翻译不加任何说明,对意义相近但语体不同的词,做同样的翻译处理,这也会让学习者产生疑惑。

汉语中具有典型语体特征的词汇出现的形式是有限制的。对于只用于书面语的语言点,要尽量避免采用对话的形式,而应该多采用转述类的练习形式。在编写课后练习时,也要考虑练习所设定的语境,练习的语言点与练习形式在语体上应该互相照应,并提示学生注意一些语体词汇的使用条件。

注释

1. 国家语言资源监测与研究有声媒体中心成立于2005年9月16日,是教育部语言文字信息管理司和中国传媒大学共建的研究中心。
2. 以每学年40周,每周24学时计算。

基金信息

本研究系国家社科基金一般项目(项目编号:21BYY170)、国务院侨办2022年度华文教育研究课题(项目编号:22GQB181)、世界汉语教学学会全球中文教育主题学术活动计划(项目编号:SH22Y17)、同济大学语言文字推广基地双强项目(项目编号:TJSQ22ZD03)的阶段性成果。

参考文献

- Biber,D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Cobb,T.(2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review*, 59(3):393-423.
- Petch-Tyson,S. (1998). Writer/Reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp.107-118). Longman.

- S·皮特·科德(1983)《应用语言学导论(中译本)》,上海外语教育出版社。
- 仓石武四郎(1963)《中国语辞典》,岩波书店。
- 曹炜(2003)现代汉语口语词和书面语词的差异初探,《语言教学与研究》,6:39-44。
- 冯胜利(2006)《汉语书面用语初编》,北京语言大学出版社。
- 冯胜利(2007)《汉语综合教程》,高等教育出版社。
- 冯胜利、王洁、黄梅(2008)汉语书面语体庄雅度的自动测量,《语言科学》,2:113-126。
- 冯胜利(2010)论语体的机制及其语法属性,《中国语文》,5:400-412+479。
- 冯胜利(2018)《汉语语体语法概论》,北京语言大学出版社。
- 冯胜利、王永娜、王丽娟(2021)汉语单双音节对应词的语体对立及语法属性,《汉语学报》,2:2-17。
- 郭锐(2002)《现代汉语词类研究》,商务印书馆。
- 胡裕树(1981)《现代汉语》(增订本),上海教育出版社。
- 黄伯荣、李炜(2016)《现代汉语》(第二版),北京大学出版社。
- 汲传波(2009)中级综合汉语教材语体不对应研究,《云南师范大学学报(对外汉语教学版)》,6:9-14。
- 李泉(2003)基于语体的对外汉语教学语法体系构建,《汉语学习》,3:49-55。
- 李晓琪(2013)《博雅汉语》,北京大学出版社。
- 马明艳(2017)汉语学习者书面语作文“口语化”倾向的语体表征,《汉语学习》,2:81-90。
- 潘璠(2012)中国非英语专业本科生和研究生书面语体的多特征多维度调查,《外语教学与研究》,2:220-232+320。
- 宋婧婧(2013)汉语口语与书面语词汇使用对比分析——基于传媒语料库,《厦门理工学院学报》,9:88-92。
- 王安节、鲍海涛(2003)《中华书面语词典》,吉林文史出版社。
- 王永娜(2012)书面语体“和”字动词性并列结构的构成机制,《世界汉语教学》,2:188-197。
- 王用源、陈宇豪(2023)语体语法视角下语体正式度的自动测量,《语言教学与研究》,4:79-90。
- 文秋芳、丁言仁、王文字(2003)中国大学生英语书面语中的口语化倾向——高水平英语学习者语料对比分析,《外语教学与研究》,4:268-274+321。
- 文秋芳(2009)学习者英语语体特征变化的研究,《外国语》,4:2-10。
- 袁晖、李熙宗(2005)《汉语语体概论》,商务印书馆。
- 张弓(1963)《现代汉语修辞学》,天津人民出版社。
- 张静(1980)《新编现代汉语》,上海教育出版社。
- 张世涛、刘若云(2002)《初级汉语阅读教程I》,北京大学出版社。
- 中国社会科学院语言研究所词典编辑室(2012)《现代汉语词典(第六版)》,商务印书馆。

投稿:2023年5月30日;接受:2023年11月10日;出版:2024年4月7日

作者简介

张江丽,博士,现任北京华文学院教授,北京华文学院华文教育发展研究中心副主任。主要研究方向:汉语第二语言学习者词汇习得及教学,东南亚地区的华文教育。主持国家社科基金2项,省部级项目5项,在《世界汉语教学》《语言文字应用》等相关专业期刊上发表论文近40篇,出版学术专著3部。

A Multidimensional Contrastive Study of the Written Features of Vocabularies Used by Intermediate and Advanced Chinese Second Language Learners

Jiangli Zhang

Beijing Chinese Language and Culture College, China

Abstract

Based on two self-built corpora, namely “The Chinese Pen Corpus for Foreign Students” and “The Chinese Pen Corpus for Native Speaker”, this paper investigates the written features of pen vocabulary of intermediate and advanced Chinese second language learners from three dimensions: fully written vocabulary, first/second personal pronoun, and modal particles, and compares them with native Chinese learners. The results show that (1) the written features of pen vocabulary of Chinese second language learners are far inferior to that of native Chinese learners; (2) that advanced Chinese second language learners have a higher degree of written features of pen vocabulary than intermediate learners; (3) that compared with intermediate Chinese second language learners, the number of fully written vocabulary used by advanced Chinese second language learners has increased significantly, but the number of first/second person pronouns and modal particles used by advanced Chinese second language learners has slightly decreased.

Keywords

Pen vocabulary, written features, corpus, native Chinese learners, Chinese second language learners

Dr. Jiangli Zhang is now a professor in Beijing Chinese Language and Culture College, the deputy director of the Chinese Education Development Research Center in Beijing Chinese Language and Culture College. Her research interests include vocabulary acquisition and teaching for Chinese second language learners, Chinese language education in Southeast Asia. She has led 2 National Social Science Foundation projects, 5 provincial and ministerial level projects, published nearly 40 papers in Chinese Teaching in the World, Applied Linguistics and related professional journals, and published 3 academic books.

References

- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Cao, Wei (曹炜). (2003). 现代汉语口语词和书面语词的差异初探 [A Preliminary Study on the Differences between the Colloquial and the Literary in Modern Chinese]. *语言教学与研究* [Language Teaching and Linguistic Studies], 6, 39-44.
- Cobb, T. (2003). Analyzing late interlanguage with learner corpora: Quebec replications of three European studies. *Canadian Modern Language Review*, 59(3):393-423.
- Corder S.P. (1983). 应用语言学导论 (中译本) [Introducing Applied Linguistics (Chinese version)], 上海外语教育出版社 [Shanghai Foreign Language Education Press].
- Feng, Shengli (冯胜利), Wang, Jie (王洁), & Huang, Mei (黄梅). (2008). 汉语书面语体庄雅度的自动测量 [An Automatic Feature Checking Algorithm for Degree of Formalities in Written Chinese]. *语言科学* [Linguistic Sciences], 2, 113-126.
- Feng, Shengli (冯胜利), Wang, Yongna (王永娜), & Wang, Lijuan (王丽娟). (2021). 汉语单双音节对应词的语体对立及语法属性 [Register Contrasts among Monosyllabic vs. Disyllabic Synonyms and Their Grammatical Properties in Mandarin Chinese]. *汉语学报* [Chinese Linguistics], 2, 2-17.
- Feng, Shengli (冯胜利). (2006). 汉语书面用语初编 [Expressions of Written Chinese], 北京语言大学出版社 [Beijing Language and Culture University Press].
- Feng, Shengli (冯胜利). (2007). 汉语综合教程 [Comprehensive Chinese], 高等教育出版社 [Higher Education Press].
- Feng, Shengli (冯胜利). (2010). 论语体的机制及其语法属性 [On mechanisms of Register System and its grammatical property]. *中国语文* [Studies of the Chinese Language], 5, 400-412+479.
- Feng, Shengli (冯胜利). (2018). 汉语语体语法概论 [An Introduction to Register-Based Grammar of Chinese], 北京语言大学出版社 [Beijing Language and Culture University Press].
- Guo, Rui (郭锐). (2002). 现代汉语词类研究 [Studies in Word Classes of Modern Chinese], 商务印书馆 [Commercial Press].
- Hu, Yushu (胡裕树). (1981). 现代汉语(增订本) [Modern Chinese (Revised and Expanded Edition)], 上海教育出版社 [Shanghai Education Publishing House].
- Huang, Borong (黄伯荣), & Li, Wei (李炜). (2016). 现代汉语 (第二版) [Modern Chinese (2nd ed.)], 北京大学出版社 [Peking University Press].
- Institute of Linguistics, Cass (中国社会科学院语言研究所词典编辑室). (2012). 现代汉语词典 (第六版) [The Contemporary Chinese Dictionary (6th ed.)], 商务印书馆 [Commercial Press].
- Ji, Chuanbo (汲传波). (2009). 中级综合汉语教材语体不对应研究 [A Study of the Stylistic Contradictions in the Intermediate Chinese Textbooks]. *云南师范大学学报 (对外汉语教学版)*. [Journal of Yunnan Normal University (Teaching & Studying Chinese as a Foreign Language Edition)], 6: 9-14. <https://doi.org/10.16802/j.cnki.ynsddw.2009.06.002>
- Kuraishi, Takejirō (仓石武四郎) (1963). 中国语辞典 [A Japanese-Chinese Dictionary], 岩波书店 [Iwanami Shoten Publishers].
- Li, Quan (李泉). (2003). 基于语体的对外汉语教学语法体系构建 [Building of Grammar System for Teaching Chinese as a Second Language Based on Style of Writing]. *汉语学习* [Chinese Language Learning], 3, 49-55.
- Li, Xiaoqi (李晓琪). (2013). 博雅汉语 [BOYA CHINESE], 北京大学出版社 [Peking University Press].
- Ma, Mingyan (马明艳). (2017). 汉语学习者书面语作文“口语化”倾向的语体表征 [Features of Oral Style in Chinese Compositions Written by CSL Learners]. *汉语学习* [Chinese Language Learning], 2, 81-90.
- Pan, Fan (潘璠). (2012). 中国非英语专业本科生和研究生书面语体的多特征多维度调查 [MF and MD analysis of written texts produced by Chinese non-English major undergraduates and graduates]. *外语教学与研究* [Foreign Language Teaching and Research], 2, 220-232+320.

- Petch-Tyson, S. (1998). Writer/Reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp.107-118). Longman.
- Song, Jingjing (宋婧婧). (2013). 汉语口语与书面语词汇使用对比分析——基于传媒语料库 [A Media-Corpus-Based Comparative Study on the Use of Vocabulary in Spoken and Written Chinese]. 厦门理工学院学报 [Journal of Xiamen University of Technology], 9, 88-92. <https://doi.org/10.19697/j.cnki.1673-4432.2013.03.021>
- Wang, Anjie(王安节), & Bao, Haitao (鲍海涛). (2003). 中华书面语词典 [Dictionary of Chinese Literary Register], 吉林文史出版社 [Jilin Wenshi Press].
- Wang, Yongna (王永娜). (2012). 书面语体“和”字动词性并列结构的构成机制 [The Structural Mechanism of Predicative Coordinating Verbal Phrases Using the Conjunction *hé* in Written Chinese]. 世界汉语教学 [Chinese Teaching in the World], 2, 188-197. <https://doi.org/10.13724/j.cnki.ctiw.2012.02.007>
- Wang, Yongyuan (王用源), & Chen, Yuhao (陈宇豪). (2023). 语体语法视角下语体正式度的自动测量 [The automatic measurement of the stylistic-register formality from the perspective of stylistic-register grammar]. 语言教学与研究 [Language Teaching and Linguistic Studies], 4, 79-90.
- Wen, Qiufang (文秋芳), Ding, Yanren (丁言仁), & Wang, Wenyu (王文宇). (2003). 中国大学生英语书面语中的口语化倾向——高水平英语学习者语料对比分析 [Features of oral style in English compositions of advanced Chinese EFL learners: An exploratory study by contrastive learner corpus analysis]. 外语教学与研究 [Foreign Language Teaching and Research], 4, 268-274+321.
- Wen, Qiufang (文秋芳). (2009). 学习者英语语体特征变化的研究 [A Study on Changes of Register Features in Learner English]. 外国语 [Journal of Foreign Languages], 4, 2-10.
- Yuan, Hui (袁晖), & Li, Xizong (李熙宗). (2005). 汉语语体概论 [An Introduction to Chinese Register Variation], 商务印书馆 [Commercial Press].
- Zhang, Gong (张弓). (1963). 现代汉语修辞学 [Rhetoric of Modern Chinese], 天津人民出版社 [Tianjin People's Publishing House].
- Zhang, Jing (张静). (1980). 新编现代汉语 [Modern Chinese: A New Compendium], 上海教育出版社 [Shanghai Education Publishing House].
- Zhang, Shitao (张世涛), & Liu, Ruoyun (刘若云). (2002). 初级汉语阅读教程 I [Elementary Chinese Reading Course I], 北京大学出版社 [Peking University Press].