

留学生汉语书面语中的口语化倾向研究*

汲传波 刘芳芳

提要 本文通过比较 37 个汉语口语格式在 HSK 动态作文语料库和现代汉语平衡语料库中使用频率的差异,证明外国留学生书面语中确实存在口语化倾向:与汉语母语者相比,留学生在书面语中过多使用其中的 19 个汉语口语格式。本文还对不同国别留学生在书面语中使用口语格式的频率进行了排序。

关键词 留学生;汉语书面语;口语化倾向

一 引言

许多学者论述过语体在对外汉语教学中的重要作用,如盛炎(1994)、丁金国(1997、1999)、李泉(2003、2004)等,但对外汉语教学界对语体关注较少(韩莹 2008)。近些年来,在汉语国际传播的大背景下,如何加强语体教学,提高学生得体交际的能力,正成为学界的讨论热点。截至目前,与留学生语体习得有关的研究成果集中在三个方面:第一,尝试利用留学生作文、汉语教材、调查问卷等了解留学生的汉语书面语体、口语语体能力,发现留学生语体能力的不足(韩莹 2008;胡晓慧 2008;刘圣心 2008;张春玲 2008;张憬霞 2009);第二,针对留学生语体能力差的情况,从教材编写、词汇分级、练习设置等方面提出建议(王晓娜 2003;汲传波 2009;曾毅平 2009);第三,从国别的视角关注语体习得问题,如周芸等(2010)、周芸等(2011)。

比较起来,国内英语教学界对语体习得的研究时间更早,方法也更科学。如程雨民(1989)就提出区别英语书面语和口语的语体成分表,彭宣维(2000)进行了补充。文秋芳等(2003)、文秋芳(2009)、潘璠(2012)利用中介语语料库与英语本族语者语料库进行对比研究,发现中国英语学习者作文中存在明显的“口语化”倾向。对外汉语教学界的研究则刚刚起步,研究方法还局限于程雨民(1989)所提到的语篇语域量化法,如张憬霞(2009)。而语料库的对比研究尚未进行。

英语教学界的研究成果值得汉语教学界参考,但因目前汉语学界没有研制出能够区别汉语口语和书面语的语体成分表,所以本文拟另辟蹊径,通过筛选一定数量的汉语常用口语格式,在 HSK 动态作文语料库中进行穷尽检索,并将检索结果与国家语委平衡语料库的结果对比,验证留学生汉语书面语中是否具有口语化倾向,以及不同国别的留学生汉语书面语中口语格式使用是否存在差异。本研究对进一步研制区别汉语书面语与口语的语体成分表有参考价值。

* 本文系北京大学对外汉语教育学院 2012 年科研项目“外国留学生汉语语体习得研究”的部分成果。

二 研究设计

2.1 语料来源

本研究语料一部分源于北京语言大学“HSK 动态作文语料库”，语料总数 11569 篇，424 万字；一部分源于国家语委现代汉语平衡语料库，约 2000 万字，该语料库的语料多样、均衡，能够作为汉语母语者使用书面语的典型代表。

2.2 口语格式的确定

我们以《汉语口语常用句式例解》(刘德联、刘晓雨 2005, 北京大学出版社)和《汉语口语词典》(施光亨 2012, 商务印书馆)作为选择常用口语格式的依据。这两本词典的作者都是具有丰富对外汉语教学经验的专家, 所编选的口语格式(词、句式)具有一定的科学性。为了缩小研究范围, 选择更恰当的研究对象, 我们以这两本词典都收入的 61 个口语格式在中介语料库中进行初步检索, 排除没有检索到的格式, 共确定 37 个口语格式为本文的考察对象(见表 2)。

2.3 语料分析说明

一般认为典型的口语是面对面的对话, 典型的书面语是议论文。我们虽然把语料库中的作文都视为留学生书面语语料, 但是为了区分典型和非典型, 又通过分析作文题目将语料分为记叙文体和议论文体两大类。记叙文体是非典型的书面语, 议论文体是典型的书面语。记人、叙事类的作文归入记叙文体, 比如“记对我影响最大的一个人”“一封写给父母的信”“我的城市/乡村生活”“记我的父亲”“我的童年”“我的一个假期”等。表达观点类的作文归入议论文体, 比如“如何看待‘安乐死’”“静音环境对人体的危害”“由‘三个和尚没水喝’想到的”“吸烟对个人健康和公众利益的影响”“如何解决代沟问题”“父母是孩子的第一任老师”“学习汉语的苦与乐”“我对男女分班的看法”“我看流行歌曲”“绿色食品与饥饿”等。

我们对 37 个口语格式逐一在中介语料库和平衡语料库中检索, 得到使用总次数。然后根据各语料库的规模, 计算出各口语格式的标准化频率均值(每百万字)。本文所有表格中的标准化频率均值在计算时都为“出现总次数/每百万字”, 不再逐一标注。另外, 我们还对这些口语格式在中介语料库中的记叙和议论两种文体中的使用频率分别进行了统计, 对不同国别的留学生所使用的口语格式频率也进行了统计。

需要说明的是, 因为某些格式有不同的语义, 在统计时按照《汉语口语常用句式例解》和《汉语口语词典》的解释为准。比如, 对“好了”这一格式, 只统计其在句首位置、独立作为小句使用的例句; 对“要说”这一格式, 也只统计其位于句首时的用例; 对“没什么”这一格式, 只统计其单用的用例, 不统计诸如“没什么钱”之类的用法。

三 分析与讨论

3.1 口语格式在两大语料库中的频率均值比较

通过统计 37 个口语格式在中介语料库和平衡语料库中的使用总次数, 我们得出了标准化频率均值(见表 1)。

表 1 37 个口语格式在语料库中的频率均值比较

中介语料库频率均值		平衡语料库频率均值
记叙文体(436.55)	议论文体(202.84)	128.55
258.02		

如表 1 所示,37 个口语格式在中介语中出现的频率均值为 258.02,要远远高于这些格式在平衡语料库中的频率均值 128.55。即使只看 37 个口语格式在中介语议论文体(典型的书面语体)中的频率均值 202.84,也要高于其在平衡语料库中的统计数据。这些数据说明,外国留学生书面语中使用口语格式的频率要高于汉语母语者。据此可以证实外国留学生书面语中确实有口语化倾向。虽然研究方法不同,但是我们的这一发现进一步印证了刘圣心(2008)、韩莹(2008)、张憬霞(2009)等的研究结论。同时,这一研究结果也与英语作为第二语言教学界的发现一致(文秋芳等 2003;文秋芳 2009;潘璠 2012),即英语作为外语学习的学习者书面语中有口语化倾向。

为什么高级水平的留学生在汉语书面语中会有口语化倾向,原因不难解释。首先是目前对外汉语教学界对语体的重视程度不够,在教材编写、课堂教学、测试等诸多环节都很少涉及语体,这必然导致高级阶段的留学生不能真正地区分汉语书面语和口语,导致留学生语体意识薄弱。另外,还可能因为提高留学生的语体意识非常有难度,教师即使想重视也往往不得要领,不知如何进行教学才更有效。有学者就曾敏锐地指出,语体习得与语言习得的显著不同点是:语言习得大致有个时限,并且可以规定出一个习得的标准,即语言能力的水平等级;而语体意识的习得则不然,既无时限,也很难找出一个共同的标准(丁金国 1999)。既然如此,高级阶段的留学生在语体能力方面的欠缺似乎也是必然。这说明,高级阶段的外国留学生对于汉语的学习也还远未结束,即使作文成绩取得高分,也需要在语体方面继续努力。我们在对海外本土中文教师的培训中也发现,有些教师虽有多年中文教学经验,汉语口语水平也非常高,但仍然对汉语语体的区分不甚清楚。

3.2 具体情况分析

3.2.1 外国留学生使用口语格式的具体情况

表 2 口语格式在中介语中的频率均值排序

排序	格式名称	频率均值	排序	格式名称	频率均值
1	不好意思	50.94	20	巴不得	1.42
2	就这样	34.91	20	怪不得	1.42
3	无所谓	20.99	22	(就)得了	1.18
4	没问题	20.75	22	你说呢	1.18
5	不得了	19.10	22	说真的	1.18
5	不用说	19.10	25	没劲	0.94
7	不怎么(样)	15.57	26	没什么	0.71
8	对了	14.39	26	有的是	0.71
9	好了	8.49	26	怎么搞的	0.71
10	你看	8.02	29	话是这么说	0.47
11	罢了	6.37	30	说不好	0.47
11	话说回来	6.37	30	说了算	0.47
13	(不)要说	4.95	30	要我说	0.47
14	算了	4.25	30	一句话	0.47
15	就是了	3.77	34	本来嘛	0.24
16	不像话	2.12	34	别看	0.24
17	不要紧	1.89	34	够朋友	0.24
18	没事儿	1.65	34	说白了	0.24
18	再说吧	1.65			

表 2 显示,外国留学生在书面语中对 37 个口语格式使用最多的是“不好意思(50.94)、就这样”(34.91)等,而使用最低的为“本来嘛 (0.24)、别看(0.24)、够朋友(0.24)、说白了(0.24)”,最高与最低的使用频率之比高达 212.25:1。

产生以上结果的原因有四点:第一,外国留学生对于汉语口语格式掌握的熟练程度不同。对“不好意思、就这样”等掌握得比较熟练,而对“够朋友、说白了”等不甚熟悉;第二,对外汉语教学实践中语体教学薄弱导致的留学生语体意识薄弱。留学生掌握了这些汉语口语格式,但是不清楚其语体使用特点,导致其在书面语中使用过多。第三,可能是因为作文话题的限制,导致留学生对部分口语格式使用较少;第四,这些口语格式本身也有使用频率的高低差异。

3.2.2 留学生在书面语中过多使用的口语格式

只比较 37 个口语格式在中介语中频率均值的差异,无法确切地知道到底哪些口语格式被留学生在书面语中过多使用。为此,我们把 37 个口语格式在中介语中的标准化频率均值与其在现代汉语平衡语料库中的标准化频率均值进行了比较,抽取出来比较结果为正数的部分并按照频率均值差从高到低进行排序,结果见表 3。

举例来说,比如“不得了”这一口语格式在中介语中的使用频率均值为 19.1,在现代汉语平衡语料库中的使用频率为 3.55,二者之间的差为 15.55。这一结果为正数,说明留学生在书面语中使用“不得了”这一口语格式的频率要高于汉语母语者。

表 3 口语格式在中介语与平衡语料库中的频率均值比较

序号	格式名称	均值差	序号	格式名称	均值差
1	不好意思	41.54	11	不像话	1.67
2	没问题	18.5	12	再说吧	1.65
3	无所谓	15.74	13	没事儿	1.55
4	不得了	15.55	14	话是这么说	0.37
5	不用说	14.2	14	要我说	0.37
6	对了	13.79	16	(就)得了	0.33
7	不怎么(样)	13.37	17	说不好	0.27
8	就这样	12.36	18	本来嘛	0.09
9	好了	6.14	18	没劲	0.09
10	话说回来	4.72			

表 3 说明, 37 个口语格式中,有 19 个口语格式在外国留学生书面语中使用的频率要高于汉语母语者,另有 18 个口语格式外国留学生书面语中的使用频率低于汉语母语者。另外一个发现是,这 19 个口语格式的使用频率均值差异是极不均衡的。最高的“不好意思”留学生和汉语母语者频率均值差异达到 41.54;有 7 个口语格式的频率均值差异在 10-20 之间,如“没问题、无所谓、不得了”等;有 5 个口语格式的频率均值差异在 1-7 之间,如“好了、话说回来、不像话、再说吧”等;还有 6 个口语格式的频率均值差异小于 1,比如“本来嘛、没劲”等。对比表 2 可以看出,频率均值差异排序靠前的口语格式,如“不好意思(41.54)、没问题(18.5)、无所谓(15.74)、不得了(15.55)、不用说(14.2)、对了(13.79)、不怎么样(13.27)、就这样(12.36)”等,也是留学生在书面语中使用频率最高的,被留学生在书面语中使用的也最多。

为了进一步了解留学生在严格的书面语体中对口语格式的使用情况,我们又计算了留学生在议论文体中使用这些口语格式的频率均值,并将之与现代汉语平衡语料库的统计结果进行比较(见表 4)。

表 4 口语格式在中介语议论文体与平衡语料库中的频率均值比较

序号	格式名称	均值差	序号	格式名称	均值差
1	没问题	17.82	10	对了	2.18
2	不好意思	17.77	11	再说吧	0.93
3	无所谓	17.60	12	(就)得了	0.69
4	不用说	14.86	13	话是这么说	0.52
5	不怎么(样)	12.31	14	没事儿	0.21
6	不得了	9.42	14	要我说	0.21
7	就这样	7.40	16	本来嘛	0.16
8	话说回来	5.76	17	说不好	0.11
9	不像话	2.33	18	就是了	0.01

表 4 说明,在典型的书面语体中,外国留学生仍过多使用一些口语格式。“没问题、不好意思、无所谓、不用说、不怎么(样)、不得了、就这样、话说回来、不像话、对了”等 10 个格式的使用频率均值都比汉语母语者明显偏高。这进一步证明,留学生书面语中的口语化倾向确实存在。

四 不同国别的留学生书面语中的口语格式使用情况

为了了解不同国别的留学生在汉语口语格式使用上的差异,以便能够进行有针对性的语体教学,我们对这 37 个口语格式的使用情况进行了分国别统计。使用 37 个口语格式的国家共有 41 个,其中 13 个国家的留学生仅使用一次,它们是葡萄牙、塔吉克斯坦、奥地利、希腊、约旦、罗马尼亚、瑞士、芬兰、保加利亚、土库曼斯坦、意大利、秘鲁和西班牙。考虑到这种情况,本文只对大于和等于 5 次的进行统计,共有 13 个国家(见表 5)。

HSK 动态作文语料库中没有各国留学生作文的字数说明,需要通过检索各国留学生作文篇数进行计算。我们先检索到要统计的 13 个国家的留学生作文篇数,除以总篇数 11569,再乘以 424 万字,得出各国留学生作文字数。然后用口语格式的出现次数除以各国留学生的作文字数,得出频率均值。比如,韩国留学生在 HSK 动态作文语料库中的篇数最多(4171 篇),在语料库中占 36.05%。因为语料总字数为 424 万字,所以根据韩国学生的作文所占比例可以大致推算出韩国学生的语料总数为 152.87 万字。因为韩国学生使用 37 个口语格式的总次数为 454 次,所以其使用频率均值为 297(每百万字)。

表 5 不同国家留学生口语格式使用比较

	柬埔寨	蒙古	德国	韩国	印尼	日本	缅甸	英国	越南	新加坡	泰国	马来西亚	澳大利亚
篇数	23	59	43	4171	739	3211	202	108	221	843	374	422	123
次数	6	15	6	454	77	309	19	10	18	63	26	29	7
频率	712	694	381	297	284	263	257	253	222	204	190	188	155

表 5 说明,不同国家的留学生汉语书面语中的口语化倾向存在差异,在这 13 个国家中,柬埔寨学生使用口语格式的频率最高,澳大利亚学生最低。本研究结论和英语教学界的研究成果相似,比如文秋芳等(2003)的研究也发现不同母语背景的英语学习者书面语中口语化倾向的程度并不完全一样。

表 5 中越靠前的国家,该国留学生在汉语书面语中越多地使用口语格式,更多地表现出口语的倾向性。柬埔寨、蒙古、德国等国家留学生的书面语中存在的口语化倾向值得特别关注。当然这一结论还需在更多的研究中得到证实。另外,需要说明的是,本研究仅是从一个维度去

观察留学生的口语化倾向所作出的判断,不排除运用多维分析法之后,不同国家留学生口语化倾向的排序有所变化。

另外,从表5的统计也可以看出,虽然韩国、日本的留学生在 HSK 作文中所占比例最高,使用汉语口语格式的次數也最多,但是相对于其他国家来说,使用汉语口语格式的频率并不属于最高的。韩国排第4位,日本排第6位。文秋芳等(2003)认为造成英语作为第二语言学习者书面语中口语化倾向存在差异的原因可能有多种,一个原因可能是由英语水平的差异所造成,另一个重要原因可能是教学方法的不同导致了不同的学习效果。从韩、日留学生的情况来看,我们认为造成差异的原因更可能是这两个国家学生的学习策略、性格特点的不同。在教学实践中我们会发现,日本留学生往往比韩国学生更不爱开口说话,这可能是导致日本学生比韩国学生口语格式使用频率低的一个重要原因。当然,也有可能是在 HSK 语料库中韩国学习者的汉语水平整体略低于日本学习者,从而造成韩国学习者的口语化倾向更明显。这有待于将来的进一步研究来证实。

五 研究启示

本文从外国留学生在汉语书面语中使用口语格式这一视角进行研究,发现留学生书面语中确实存在口语化倾向。与汉语母语者相比,留学生过多使用一些汉语口语格式。这告诉我们:在汉语口语格式的教学中,除了教会留学生掌握其语法、语义、语用特点,更要教会留学生掌握其语体特点。英语教学界的观点值得我们借鉴:教师需帮助学习者识别正式的书面语与口语之间的差异,注重功能与语言实现手段之间的联系,教授如何正确使用词汇和语法特征构建符合语体要求的语篇(潘璠 2012)。

目前来看,对外汉语教学界对于语体的重视程度还远远不够。高级水平的留学生虽然也在学习精读、口语、写作等课程,但是对于语体的系统训练仍然偏少,很多留学生的语体转换能力非常薄弱。有过外国研究生论文指导经验的教师都有深切的体会,虽然有些研究生的口语表达能力已经很好,但是他们的论文语言仍然呈现口语化的特点(当然也有语法问题),修改起来非常费时费力。因此,我们同意丁金国(1997)的看法:为了有效地提高对外汉语教学的水平,必须将“语体意识”的培育置于对外汉语教学的核心地位。

要达到语体教学的目标,只有形而上的“语体意识”培育的指导思想还不够,因为留学生的语体习得不能没有时限,不能没有具体标准。如果没有标准和时限,汉语教学大纲、教材编写、课堂教学实践等都无据可依。目前的语体教学现状就是如此。虽然很多学者都提出要重视对外汉语语体教学,但是缺乏明确、具体的标准,使得一线对外汉语教师不知道该如何操作。因此,我们建议语体习得设一个低、中、高的标准,并区分“语体理解”与“语体表达”两个层面。对留学生来说,语体表达比语体理解要困难的多,这也理应成为语体教学中最应该关注的部分。

要解决以上难题,目前汉语教学界迫切需要研制出能够区分汉语口语和书面语的语体成分表。英语教学界早有学者对此进行专门研究,比如 Biber(1988)就对英语口语、书面语的大量语言特征进行了多维度分析,系统揭示了各类口语、书面语的语域间差异。在汉语教学界,冯胜利(2003、2006、2008)也曾就汉语书面正式语体的特征、教学、书面语体庄雅度的自动测量等进行了一系列探讨,颇有影响。虽然如此,汉语学界关于区分口语、书面语语体特征方面的研究成果仍然偏少,关于区分口语、书面语的语体成分表的研制始终没有取得重要突破,这都影响和制约着对外汉语语体教学实践。我们期待有更多的学者关注并研究这一课题。

参考文献

- 程雨民 1989 《英语语体学》，上海外语教育出版社。
- 丁金国 1997 对外汉语教学中的语体意识，《烟台大学学报(哲社版)》第 1 期。
- 丁金国 1999 再论对外汉语教学中的语体意识，《语言文字应用》第 2 期。
- 冯胜利 2003 书面语语法及教学的相对独立性，《语言教学与研究》第 2 期。
- 冯胜利 2006 论汉语书面正式语体的特征与教学，《世界汉语教学》第 4 期。
- 冯胜利 2008 汉语书面语体庄雅度的自动测量，《语言科学》第 2 期。
- 韩莹 2008 中高级对外汉语综合课教材中书面语体情况考察与分析，北京语言大学硕士学位论文。
- 胡晓慧 2008 试析留学生汉语写作中的口语体倾向，《华侨大学学报(哲社版)》第 3 期。
- 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)》第 6 期。
- 李泉 2003 基于语体的对外汉语教学语法体系构建，《汉语学习》第 3 期。
- 李泉 2004 面向对外汉语教学的语体研究的范围和内容，《汉语学习》第 1 期。
- 刘圣心 2008 高级阶段留学生书面语体意识的培养与考察，暨南大学硕士学位论文。
- 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》第 2 期。
- 彭宣维 2000 《英汉语篇综合对比》，上海外语教育出版社。
- 盛炎 1994 跨文化交际中的语体学问题，《语言教学与研究》第 2 期。
- 王晓娜 2003 第二语言语体能力的培养与教材虚拟语境的设置，《汉语学习》第 1 期。
- 文秋芳 2009 学习者英语语体特征变化的研究，《外国语》第 4 期。
- 文秋芳、丁言仁、王文宇 2003 中国大学生英语书面语中的口语化倾向，《外语教学与研究》第 4 期。
- 曾毅平 2009 语体理论在对外汉语教学中的应用，《修辞学习》第 5 期。
- 张春玲 2008 中高级阶段越南学生汉语语体偏误分析，广西民族大学硕士学位论文。
- 张憬霞 2009 高等 HSK 考试中中学生书面语能力考察与分析，北京语言大学硕士学位论文。
- 周芸、张婧 2010 泰国学生汉语谈话语体能力习得调查，《云南师范大学学报(对外汉语教学版)》第 3 期。
- 周芸、张永芹、张婧 2011 论泰国学生汉语语体能力培养的基本原则，《云南师范大学学报(对外汉语教学版)》第 2 期。
- Biber, Douglas 1988 *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

Colloquial Trends in Chinese Compositions of Foreign Students

Abstract This study testifies the colloquial trends which can be found in the writings of foreign students by comparing the difference of frequencies of 37 spoken Chinese formats from the interlanguage corpus and the modern Chinese corpus. It is found that 19 spoken Chinese formats are overused by foreign students comparing with the native Chinese speakers. This study also sequences the frequencies of 37 spoken Chinese formats used by foreign students from different countries. Based on these findings, the strategies about how to teach Chinese to foreign students from a stylistic perspective is discussed.

Key words foreign students; writing Chinese; colloquial trends

(汲传波 100871 北京大学对外汉语教育学院)

(刘芳芳 100089 北京外国语大学中文学院)

(责任编辑 李先银)