

# 中国大学英语学习者英语书面语中的口语化倾向

郟阳医学院 刘兴兵

**摘要:** 本研究使用作者/读者显现度和不同词频等级的词汇分布两个参数,运用对比分析语料库的方法,回答三个问题:1. 中国大学英语学习者的书面语中是否存在明显的口语化倾向;2. 口语化倾向是否随着英语水平的提高而有所弱化;3. 前两个问题的回答对大学英语教学的启示是什么。研究结果表明:中国大学英语学习者的书面语中的确存在着很强的口语化倾向;这种口语化倾向并没有预期地随着英语水平的提高有明显的弱化,作者/读者显现度各个特征的发展很不均衡;需要从母语影响、大学英语大纲、教材、教法、学习策略、中介语特点等方面进行深入探讨,采取对策,促使学生正确使用英语书面语进行写作。

正如桂诗春、杨惠中(2002)所述,对于将英语作为外语或第二语言的学习者,其英语书面语基本上就是“写话”,较接近于英语为母语的英美人的口语。多位国内外学者,如 Ringbom (1998)、Petch-Tyson (1998)、Cobb (2003)、文秋芳等(2003),都从不同的方面用语料库的方法证明了在高水平英语学习者(英语专业三、四年级学生)的书面语中确实存在着口语化的倾向,具体表现为:一是过度强化了作者/读者显现度(writer/reader visibility),二是超用2000个高频词,致使单词缺乏区分性和多样性;而且这种倾向随着高水平学习者英语水平的提高逐渐减弱。本文旨在参照作者/读者显现度和不同词频等级的词汇分布两个参数,使用中国学习者英语语料库与英语本族语者口语和书面语语料库进行对比,研究另一类型的学习者<sup>4/4</sup>中国大学英语学习者(非英语专业学生)的书面语,以验证这一结论是否具有普遍意义,考察在中国大学英语学习者的英语书面语中是否也存在着这种口语化倾向,其发展趋势怎样,以及这种倾向和发展趋势对大学英语教学有何种启示。正如桂诗春、杨惠中(2002)和 Ringbom (1998)等指出的那样,国内外高级英语学习者书面语中口语化特征的另一个特点表现在词汇上。由于学习者的词汇量小且多样性和区分性差,就导致了常用词的超用与低频词的少用。本文将在三个不同的词频等级上(见表3)比较中国大学英语学习者语料和本族语者语料,以揭示中国非英语专业学生英语书面语中的词汇的超用和少用情况及其发展。

## 一、研究设计

1. 研究问题。本项研究主要解决三个问题:

(1) 中国大学英语学习者书面语中是否存在着口语化倾向?

- a) 与本族语者的书面语和口语相比,中国大学英语学习者英语书面语中的作者/读者显现度有多高?
- b) 与本族语者的书面语和口语相比,中国大学英语学习者英语书面语中词汇在不同词频等级的分布情况怎样?

(2) 中国大学英语学习者英语书面语中的口语化倾向是否随着英语水平的提高而有变化?

- a) 参加四级和六级的学生使用突出作者/读者显现度的语言标记数量是否随着年级的提高而有所变化?
- b) 两个层次的学生作文中词汇在各个词频等级分布的情况是否随着年级的提高而有所变化?

(3) 这些研究结果对大学英语教学的启示是什么?

2. 语料库和检索软件。本项目使用的语料库主要有四个:中国学习者英语语料库(CLEC)的大学英语部分(st 3和st 4)、英国国家语料库的第一版的口语部分(BNC Spoken)、英国国家语料库世界版的口语部分(BNC World Spoken)和由Linguistic Data Consortium在线提供的Brown语料库(<http://www ldc upenn edu/cgi-bin/ldc/textcorpus?doc=yes&corpus=BROWN>)。中国学习者英语语料库的大学英语部分包括两个子库St 3和St 4,分别收集参加大学英语四级和六级的学生的英语作文,主要是议论文。后三个都是本族语语料库,其中BNC Spoken和BNC World Spoken代表典型的英语口语,Brown代表典型的英语书面语。

本文的第二个问题试图描述中国非英语专业学生书面语中口语化倾向随着英语水平的提高而发生的变化,所以首先必须确定参加六级考试的学生是否比参加四级考试的学生的英语水平有一定的提高,即比较St 4和St 3所代表的英语水平。表1提供了一些参数在两个子库和Brown中的数据。同st 3相比,st 4在平均句长和标准词型/词次比上有较大提高;st 4的平均

词长有少许降低,可能是由于 st 4 中全是试卷作文,写作时不能借助词典,而 st 3 中还存有将近一半的课堂作文和课后作文,可以查阅词典和其他资料。因此可以得出结论,这两个阶段的学生在英语水平上确实有一定的提高。

表1 St3、St4 和 Brown 的对数型/次比、平均词长和平均句长

	平均词长	平均句长	标准词型/词次比
St	34.05	17.92	31.55
St	44.02	19.31	32.75
Brown	4.47	27.84	44.55

本研究使用的检索软件有 Concapp Windows、Concordance 3.0 和 CLEC 附带的 Myfinder。

## 二、结果与讨论

1. 书面语中的口语化倾向。本项研究所要回答的第一个问题是:中国大学英语学习者英语书面语中是否存在口语化倾向,程度何如。该问题的回答涉及两个参数:(1)作者/读者显现度;(2)不同词频等级的词汇分布情况。

(1) 作者/读者显现度。表 2 列出了 st3、st4、Brown 和 BNC World Spoken 中的作者/读者显现度,为了直观,表中仅仅收入了每万词的频数,但是由于表中包括了各个语料库的总词次数,所以不难计算出各个特征的原始频数。这里我们主要将 CLEC 的整个大学英语部分(St3 + St4)同两个本族语语料库进行比较。

表2 St3、St4、St5、St6、Brown 和 BNC World Spoken 中的作者/读者显现度

语料库及其大小 作者/读者呈现度特征(/万词)	St3	St4	St3 + St4	St 5	St 6	Brown	BNC World Spoken
	20.9043 万词	21.2855 万词	42.1898 万词	21.451 万词	22.6106 万词	101.43 万词	1182.7898 万词
第一人称代词单数 I, I 'x , me , my , mine	182.8	223.3	203.3	307.4	58.9	72.9	289.4
第一人称代词复数 we ,we 'x , us , our ,ours	333.4	235.4	284.0	1099	101.0	46.2	106.0
第二人称代词单复数 you ,you 'x ,your ,yours	94.9	131.0	113.1	152.3	41.6	42.2	252.4
人称代词合计	611.1	589.7	600.4	569.5	201.5	161.4	647.8
阅读/写作情境指代 here , now , this essay	16.2	12.5	14.4	21.4	12.6	19.4	40.2
强调小品词 just , really	6.9	12.4	9.7	21.6	18.7	10.2	49.0
模糊词 kind/ sort of , and so on	13.4	16.4	14.9	10.1	6.9	3.3	13.1
情境指代/强化小品词/模糊词合计	36.5	41.3	39.0	53.1	38.2	32.9	102.3
作者/读者显现度 总体特征	647.6	631.0	639.3	622.6	239.7	194.3	750.2

总体来看,St3 + St4 中作者/读者显现度特征每万词的数量 Brown 的 3.3 倍,趋向接近 BNC World 的口语部分,达到它的 85.2%。这说明,在整体上中国大学英语学习者书面语中确实存在着严重的口语化倾向。

再具体到作者/读者显现度特征的各个组成部分。St3 + St4 中人称代词每万词的数量是 Brown 中的 3.7 倍,达到了 BNC World Spoken 的 92.7%。其中第一人称代词单数和第二人称单复数在每万词的数量分别是 Brown 中的 2.8、2.6 倍,但仍少于 BNC World Spoken 中的频率;而第一人称代词复数每万词的数量是 Brown 中的 6.1 倍,甚至超过了 BNC World Spoken,是它的 2.7 倍。这说明在中国英语学习者的书面语中,人称代词存在着极其严重的超用现象,达到了本族语者书面语的三倍以上,接近了本族语者口语中的使用频率。特别是第一人称代词复数,在中国大英学习者书面语中的使用频率分别达到了本族语书面语。

而情境指代/强化小品词/模糊词在大学英语学习者书面语中的使用情况出人意料,其每万词的总计频率在 st3 + st4 中仅是 Brown 的 1.2 倍,是 BNC World Spoken 中的 38.1%。这是否表明在书面语中中国大英学习者对情境指代/强化小品词/模糊词的使用已经接近本族语者的水平呢?这种推论显然是不正确的,因为具体到其中三个不同的成分,情况又不同:阅读/写作情景指代的使用频率仅是 Brown 的 74.2%,接近 BNC World Spoken 的 35.8%,存在着比较严重的少用现象;强调小品词频率同 Brown 接近,达到它的 95.1%,是 BNC World Spoken 的 19.8%,相对本族语者书面语存在着轻微的少用现象;模糊词的差别较大,在 st3 + st4 中是 Brown 的 4.5 倍,超用程度非常突出,甚至超过了本族语者口语 BNC World Spoken 中的水平。

从作者/读者显现度的六个特征来看,中国大学英语学习者对它们的使用相对于本族语者呈现出不平衡的状况,有些特征严重超用,有些严重少用,还有一些特征同本族语者书面语中的频率接近。超用和少用的原因怎样,频率同本族语接近的

特征在使用上是不是也相同。这些都是我们下一步需要研究的问题。例如,在情境指代/强化小品词/模糊词总体特征中,仅有模糊词严重超用,这应该是不难理解的,大学英语学习者的词汇量小,词汇的多样性、区分性较低,因而很多思想无法用精确的词句表达,只好退而求其次借助模糊词达到交际目的。

(2)不同词频等级的词汇分布。表3列出了四个语料库在四个词频等级上的分布情况,St 3和St 4中各个词频等级的分布来自于CLEC光盘中的词频排列表(按频数)。在st 3+st 4中,第一个千词的频数占了总频数的91.8%,超过Brown中第一个千词所占的比例23.3个百分点,也超过BNC口语部分10.5个百分点,说明中国大英学习者在书面语中对其最常用的一千词的超用非常严重,甚至超出了本族语者口语中第一个千词的使用程度;第二个千词占总词次数的4.5%,大大低于Brown中第二个千词的比例,同BNC的口语部分相当接近,说明相对于本族语者书面语,中国非英语专业学生在书面语中对第二个千词已经出现了比较严重的少用现象,接近本族语者口语中对第二个千词的使用频率;第三个千词在st 3+st 4中的比例不但远远低于Brown,而且还少于BNC口语部分,三个千词以外的词汇在st 3+st 4中所占的词次比例相对于本族语者语料库变得更小。这表明,由于中国大学英语学习者的词汇量较小,因而相对于本族语者,大量超用其最常用的一千个词,比高水平学习者的超用范围更窄;对于其他非常用词,随着词频的降低,少用现象越来越严重。从不同等级的词汇分布看,中国大英学习者书面语中确实存在着严重的口语化倾向。

表3 不同词频等级的词汇分布

	St3	St4	St3 + St4	Brown	BNC Spoken
第一个千词	92.8%	92.2%	91.8%	68.5%	81.3%
第二个千词	4.3%	4.4%	4.5%	7.3%	4.6%
第三个千词	1.5%	1.8%	1.7%	4.3%	2.4%
其它	1.4%	1.6%	2.0%	19.9%	11.7%

总结以上两个参数的统计结果,可以得出结论:首先,在中国大学英语学习者的书面语中存在着相当严重的口语化倾向;其次,作者/读者显现度的各个特征的使用情况相对于本族语者书面语并不均衡,有的超用,有的少用,各个特征的超用和少用程度不一,这种不均衡是我们应该特别注意的,其原因有待在后续工作中从母语影响、大纲、教材、教法、学习策略、中介语特点等方面进行研究;第三,结合Ringbom (1998)、Petch-Tyson (1998)、Cobb (2003)、文秋芳等(2003)对母语为丹麦语、芬兰语、法语、瑞典语、汉语的高水平英语学习者的研究结果,可以证明这种口语化倾向不但受文化背景的影响,而且在不同水平的中介语中都普遍存在。

2. 书面语中的口语化倾向与英语水平。本项研究要回答的二个问题是第中国大学英语学习者书面语中的口语化倾向是否随着英语水平的提高有所弱化。我们仍根据表2提供的参加四级和六级的学生的书面语的数据,从作者/读者显现度和不同词频等级的词汇分布两个方面来考察他们口语化倾向的变化。

(1)从作者/读者显现度特征的总体频率来看, st 4只比st 3减少了3.9%,差别并不明显。再具体到人称代词和情境指代/强化小品词/模糊词合计两个参数,人称代词在St 4中每万词的频率比st 3中只减少了5.4%,进步也不明显,其中第一人称代词单数和第二人称单复数反而有所增加,只有第一人称复数减少了27.1%;不论是总体还是其中的三个具体特征,人称代词在两个水平的语料库中的频率普遍比Brown多用。st 4中的情境指代/强化小品词/模糊词的总体频率反而比st 3增加了24.6%,其中强调小品词增加了将近一倍,比Brown中超用21.6%;模糊词增加了22.6%,超过了Brown中的水平;只有阅读/写作情境指代有少许降低,但相对于Brown,少用情况更加突出,仅达到其中的64.4%。这说明,随着英语水平的提高,中国大学英语学习者在作者/读者显现度特征上并没有得到长足的同步发展,其表现出来的口语化倾向并没有系统弱化,而是显示出有些特征减少、有些特征增加的不均衡发展趋势。这跟我们的预期是相悖的。其中,增加了的特征有第一人称单数、第二人称单复数、强调小品词和模糊词等,它们增加的原因值得我们研究。

我们把大学英语学习者对这些特征的习得发展情况同中国英语专业学生(st 5代表英语专业一二年级学生, st 6代表三四年级)(表2和图1)做了一个纵向比较(注意:这里只能将st 3同st 4、st 5同st 6比较, st 4和st 5代表不同质的学习者,不能互相比较),以试图解释这种不均衡发展现象。从图中可以看出,英语专业学生在各个特征的频率上大都有明显的减少。由于本族语小孩在学习母语时习得各个语言特征的先后顺序不同,即不同的特征在不同的语言水平上习得,据此我们可以推测,对于中国英语学习者来说,以上六个特征在口头语和书面语上的差别的习得也存在着先后不同的现象,有的特征在口语和书面语上的差别在大学英语水平阶段就已经开始习得,比如参加六级考试的学习者使用第一人称复数的频率就比参加四级考试的学生有所下降,所以到了英语专业阶段,下降的趋势就不太明显;其他一些特征在口头语和书面语上的区别,如第一人称单数、第二人称单复数、强调小品词和模糊词等,它们在大学英语两个学习阶段的频率仍然在增加,说明非英语专业学生到参加六级考试时还没有掌握这些特征在口头语和书面语上的差别,仍然在努力尝试和探索,只是到了英语专业高年级水平时,学习者才开始意识到这些特征在两种语域中的区别,因而频率下降的趋势才明显显现出来。当然,这只是这些特征在大学英语阶段不均衡发展的一个可能的解释,需要更多的实证研究来确认这一因素,并且找到更多的合理解释(比如阅读/写作情境指代在st 3、st 4和st 6中的频率都比Brown中低,就不能用这个原因来解释,肯定有其他的影响因素在同时作用)。

(2)从不同词频等级的词汇分布的发展来看,第一个千词在st 4中的比例比在st 3中的略小,其他词频等级的词汇在st 4中的比例都比st 3略大,但这种差别都没有超过一个百分点。这种不同词频等级词汇分布的发展趋势说明了两个事实:首先,中国大学英语学习者中参加四级考试的学生对最常用的一千个词的超用程度比参加六级的学生有所减弱,而对其他非常用词的少用程度也有所降低,因而可以得出结论,从词汇的发展来看,随着英语水平的提高,中国大学英语学习者书面语中的口语化倾向确实有所减弱;其次,这种减弱的趋势在词汇分布上显得非常缓慢,并不像我们的期望那样大。

所以,随着英语水平的提高,中国大学英语学习者书面语中的口语化倾向并没有令人满意地得到均衡的发展。在作者/读者显现度各个特征方面,学生的学习进步情况显得十分混乱,有的特征的超用情况有所缓和,更多的特征超用更加严重,这

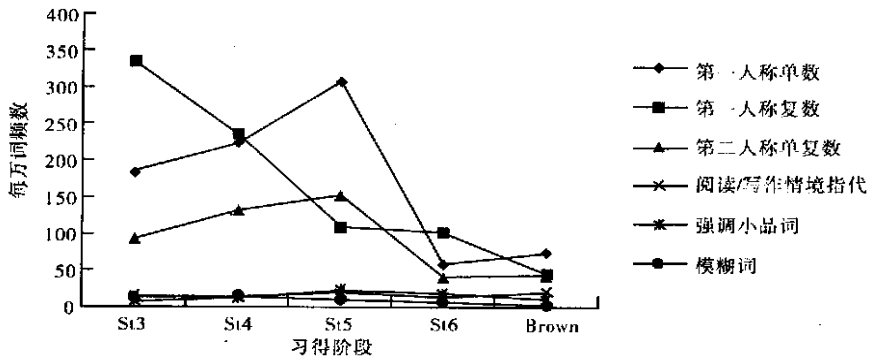


图1 作者/读者显现度各特征习得纵向比较

可能是由于中国英语学习者是在不同的学习阶段习得这些特征在口头语和书面语中的差别的,换句话说,大学英语学习者只意识到少部分作者/读者显现度特征在口头语和书面语中的差别,更多特征在两个语域中的差别只是到了英语专业阶段才被英语专业学习者所意识和学习到。从词汇等级的分布来看,口语化倾向有所减弱,但是幅度不大。这种书面语中口语化倾向同英语水平的提高不同步的现象是一个值得研究的课题,应该从母语影响、大学英语大纲、教材、教法、学习策略、中介语特点等方面去寻找原因,采取对策,以促使这种倾向能够随着学生英语水平的提高而得到削弱。

### 三、结论

本项研究所得到的主要结果归纳如下:

1. 在中国大学英语学习者的书面语中存在着很强的口语化倾向。同本族语者的书面语相比,中国学习者的各个口语化特征的严重程度并不相同,造成这种情况的原因有待研究。
2. 这种口语化倾向在不同文化背景、不同英语水平学习者的书面语中都普遍存在。
3. 与前面的几位国内外学者对高水平学习者的研究结果不同的是,我国大学英语学习者书面语中的口语化倾向并没有预期地随着英语水平的提高有明显的弱化趋势,而且作者/读者显现度各个特征的发展很不均衡,有的随着英语水平的提高而减弱,有的反而增强,笔者认为,其中的一个原因是,对这些特征在口语和书面语中的差别的习得是在不同的学习阶段或英语水平上进行的。
4. 为了解释中国大学英语学习者书面语中严重的口语化倾向以及这种倾向的发展同英语水平提高的不均衡性,我们需要从母语影响、大学英语大纲、教材、教法、学习策略、写作策略、中介语特点等方面进行深入探讨,找到原因,采取对策,以指导整个大学英语教育。首先,我们的《大学英语教学大纲》中没有明确提出书面语和口头语的分野,因而大学英语教师在教学中就很少提醒学生明确地区分两者,最终学生在写作中就没有清晰的区分两者的意识,或者学生根本就不知道两个的区别是什么,因此应该在大纲中做出明确规定,要求大学英语教师在教学中帮助学生区分语言使用在口语和书面语中的区别。其次,当代英语教学强调交互式教学,强调培养学生运用语言进行交际的能力,而口头语和书面语是交际活动中最基本的两个语域,因而我们必须在教材、教法、学习策略等方面进行改革,采取措施促使学生能够在交际中正确区分两个语域中的语言使用区别。第三,我们的大学英语教学上进行的写作练习太少,这跟大学英语教师的繁重教学任务是有关联的,如何给提供学生更多的写作练习机会,指导他们在写作中提高区分口头语和书面语的意识,这也是至关重要的。

### 参考文献

- 1 Chafe, W. L. Integration and Involvement in Speaking, Writing and Oral Literature[A]. In D. Tannen. Norwood (ed.). Spoken and Written Language: Exploring Orality and Literacy[C]. N.J.: Ablex Pub. Corp., 1982.
- 2 Cobb, T. Analyzing late interlanguage with learner corpora: Quebec replications of three European studies[J]. Canadian Modern Language Review, 2003 (3): 393 - 423.
- 3 Petch-Tyson, S. Writer/reader visibility in EFL written discourse[A]. In S. Granger (ed.). Learner English on Computer[C]. New York: Addison Wesley Longman Inc., 1998: 107 - 118.
- 4 Ringbom, H. Vocabulary frequencies in advanced learner English: A crosslinguistic approach[A]. In S. Granger (ed.). Learner English on Computer[C]. New York: Addison Wesley Longman Inc., 1998.
- 5 桂诗春,杨惠中. 中国学习者英语语料库[M]. 上海:上海外语教育出版社,2002.
- 6 文秋芳,丁言仁,王文宇. 中国大学生英语书面语中的口语化倾向[J]. 外语教学与研究, 2003 (4).