

# 基于语料库的国内英语学习者被动语态使用的对比分析

林昊<sup>1</sup>,武明飞<sup>2</sup>,冯洋<sup>1</sup>

(1. 沈阳建筑大学外国语学院, 辽宁 沈阳 110168; 2. 沈阳城市建设学院基础学科部, 辽宁 沈阳 110167)

**摘 要:**主要选用母语语料库 LOCNESS 和非母语语料库 CLEC, 应用语料库 AntConc 3.2.1w 检索软件进行检索, 利用语料库分析软件搜集、检索并进一步分析语料库数据中被动语态的使用特点。同时, 采用语言统计软件 SPSS 11.5 进行数据分析及对比, 主要运用定量分析的方式进一步对比和分析中国英语学习者议论文写作的被动语态使用特点, 总结其使用中存在的问题并分析原因, 以期为国内英语学习者的被动语态使用提供借鉴。

**关键词:**被动语态; 语料库; 二语习得; 使用特点

**中图分类号:** H313      **文献标志码:** A

被动语态不仅是英语学习中的一个非常重要而且常见的语法项目, 其使用方法及应用特点对国内学习者而言也是一个难点。在构成方式、结构特点及时态特点上, 英语和汉语的被动语态存在着巨大差异, 所以国内的英语学习者由于经常受到母语负迁移影响, 会在学习英语的过程中遇到许多困难。在国内的被动语态习得研究中, 大部分是对英语及汉语被动语态进行对比分析, 以总结两者的使用特点, 然而仍然欠缺系统运用语料库的研究手段对国内英语学习者的被动语态习得的研究。应运用不同的语料库对比及分析不同的学习者使用被动语态时表现出的具体特点, 分析、对比和总结国内英语学习者的被动语态习得情况, 进而对中国英语学习者正确运用被动语态产生的启发和促进作用。

从运用语料库对英语议论文写作中被动

语态使用情况的研究来看, 国内有很多学者已经运用语料库开展了被动语态的研究。例如: 陈万霞<sup>[1]</sup>使用国内英语学习者语料库 CLEC 研究了国内英语学习者被动语态的习得形式, 并在被动语态习得研究中发现水平不同的英语学习者对类似的被动语态的难点掌握程度也不尽相同, 同时, 被动语态使用过少是学习者容易犯的错误。但研究仅局限在被动语态的习得用法上, 并没有与本土语料库进行对比研究; 吴颖等<sup>[2]</sup>对国内英语学习者在议论文写作中被动语态的使用情况进行了研究, 并发现国内英语学习者被动语态的使用形式单一, 在高频词上存在结构固化的倾向, 在低频词上表现为先减后增的发展特点, 但并未对被动语态作出对比研究, 也没有进行错误分析研究; 刘敬伟等<sup>[3]</sup>利用 Word-smith 软件和 Claws 赋码软件对国内研究生

学位论文的被动语态使用情况进行了研究和分析,发现中国研究生学位论文被动句的使用频率高于母语者并且差异显著,但是其研究仅局限于研究生学位论文的范畴中,对更为广泛的国内英语学习者却没有展开进一步的研究;肖忠华等<sup>[4]</sup>利用语料库对国内英语学习者被动句习得个案进行了研究,从理论角度上对如何将语料库应用于教学中展开了研究和评析;王敏等<sup>[5]</sup>利用语料库对农业英语的被动语态进行了研究,发现学生大量使用被动语态的原因是为了客观表达、凸显焦点及促进语篇的衔接和连贯,但这项研究是针对农业英语的研究,没有对国内英语学习者广泛开展研究;马宇立<sup>[6]</sup>在利用语料库分析被动语态使用的错误时,指出了被动语态使用的错误主要是由于母语负迁移造成的,并给出了被动语态学习的建议,但没有在被动语态的特点上作出进一步分析。

从以上对被动语态习得的语料库研究来看,尽管当前学术界运用语料库对被动语态开展了很多研究,但均没有系统地在语料库对国内英语学习者运用被动语态使用情况上作进一步对比分析,笔者利用 Antconc3.2.1w 软件对国内英语学习者议论文写作中被动语态的使用情况展开研究,并运用数据统计软件 SPSS11.5 数据作进一步分析。通过对语料库的被动语态使用情况进一步对比分析,使英语学习者更加明晰被动语态的运用特征,正确掌握被动语态的使用特点,从而提高国内英语学习者英文写作水平。

## 一、研究对象

通过国内语料库 CLEC 和母语语料库 LOCNESS 分析国内英语学习者议论文写作中被动语态使用情况与母语者相比的区别,并探索及分析造成这种差异的原因。在原因分析基础之上,利用定量与定性相结合的方法研究国内英语学习者写作议论文时被动语态的使用特点。在这两个研究型语料库中,笔者在研究中主要利用了丛抽样(cluster sampling)的方法和软件 Antconc3.2.1w 来

搜集和整理,对语料库中的被动语态进行更为系统的研究<sup>[7]</sup>。

### 1. 国内英语学习者语料库

国内英语学习者语料库(Chinese Learner English Corpus, CLEC)是上海交通大学教授桂诗春和杨惠中共同开发、建设、组织和整理的国内英语学习者议论文语料库,也是国内对英语学习者议论文研究最为权威的语料库,搜集并整理国内学习者的作文共 100 多万词,形符数为 488 462<sup>[8]</sup>。笔者选取 CLEC 中的 ST3 和 ST4 作为对比语料库的子库,代码 ST3 指的是国内非英语专业的英语学习者参加的全国大学英语四级考试议论文,代码 ST4 指的是这些英语学习者参加全国大学英语六级考试的议论文。这两个子语料库至少在 40 万词以上,经常作为英语研究的比对话料库,笔者在对比研究中标准化处理了子语料库中的数据。

### 2. 美国母语语料库

参照语料库为美国母语语料库(The Louvain Corpus of Native English Essays Corpus, LOCNESS)。作为众多语料库研究中的权威,该语料库最终不仅确定为笔者研究的英语母语语料库,同时也作为研究的比对话料库。LOCNESS 由比利时 Louvain 大学的 Sylviane Granger 教授搜集、整理并编写,有 20 万词以上,涵盖各种文体的文章至少 200 篇,搜集了超过 182 668 形符,是当前语料库研究中经常用作对比的母语语料库。

综上所述,以上两个语料库能真实地反映英语学习者议论文的被动语态的使用情况并具有可比性。

## 二、研究方法

### 1. 确定被检索的被动句式

首先,需要确定被检索的被动句式,在英语的被动语态使用中,在本族语语料库 LOCNESS 中,经过语料库研究软件 Antconc3.2.1w 的分析可以看出,be + Ved 的结构占被动句的 87% 以上,其他表示被动的结构只占了 13%。基于以上数据比例,笔

者将着重研究被动语态 be + Ved 的结构, be + Ved 被动式不仅是英语学习者习惯运用的方式<sup>[9-10]</sup>,同时这种形式也最集中地反映了 2 个语料库中被动语态的应用特征<sup>[11-12]</sup>。其次,运用 Claws 赋码软件对 2 个语料库进行词性的赋码,便于更加清楚地分析被动语态的使用特点<sup>[13]</sup>。

2. 搜索及整理 be + Ved 型被动结构

要利用 AntConc3. 2. 1w 语料库检索软件搜索及整理出 2 个语料库中的 be + Ved 型被动结构,以便于后续的统计分析。主要针对 be 动词的现在形式 am, is, are 以及过去式 was, were, been, be 进行统一的分析研究<sup>[14-16]</sup>。

3. 统计分析

运用统计软件 SPSS 对 2 个语料库中 be + Ved 被动结构的使用频次进行对比分析,因为检验中不需要数据成正态分布,所以利用对数似然率(log - likelihood Ratio)进行差异显著性检验分析。

三、实验结果的分析

1. be 动词被动语态的整体使用特点

为了进一步研究国内英语学习者和母语者对 be 动词各种形式被动语态的使用情况,借助 Antconc 3. 2. 1w 软件对 be 动词的被动语态进行检索分析,结果如表 1 所示。

表 2 国内英语学习者和母语语料库中 be 动词的被动语态使用配对样本 T 检验

配对差异检验					T 值	自由度	显著性（双侧）
均数	标准差	标准误	95% 差异置信区间				
			低	高			
-1.965 1	7.013 0	0.665 41	-3.702 6	-0.606 9	-1.470	60	0

由表 2 可知,均数为 -1.965 1,自由度为 60,显著值为 0,小于关键值 0.05,通过数据对比可知,国内英语学习者与母语英语学习者在 be 动词引导的被动语态使用上呈显著性差异。

通过表 2 的统计结果来看,国内英语学习者在议论文写作中会有意识地运用被动语态以使文章更加正式,这种做法改变了以往国内英语学习者在议论文写作中被动语态使

表 1 国内英语学习者和母语者中被动语态 be 动词使用频次对比

be 动词	使用频次	
	CLEC 语料库	LOCNESS 母语者
am	26	1
is	562	417
are	295	193
was	192	182
were	55	91
be	687	524
been	155	138
total	1 972	1 546

表 1 的统计结果显示,国内英语学习者议论文被动语态的使用总频次为 1 972,母语者的使用频次为 1 546。通过表面的总体使用频数分析,国内英语学习者被动语态的使用频次高于母语者的使用频次,但是否具有统计学上的显著性差异,还需要进行进一步的统计分析。根据数据特点,通常需要运用统计学中的卡方检验来进行显著性检验分析。卡方检验要求语料库中的统计数据均呈正态分布,要证明数据是否呈正态分布,需要经过似然比的统计分析,经过似然比的分析得知对数似然比率为 36.76,高于  $P \leq 0.001$  时的临界值 10.83,所以,国内英语学习者议论文中的被动句的使用频次可能高于母语者并具有显著差异。但为了正确检验国内英语学习者与国外英语学习者在被动语态上是否真正具有显著性差异,还是需要将数据代入 SPSS 中的配对样本来检验(见表 2)。

用不足的情况,但因为国内英语学习者刻意追求运用被动语态这一语法形式,造成了过度使用被动语态语法结构的情况。

单独的 be 动词被动语态使用还不能说明国内英语学习者的被动语态掌握情况,还需要对被动语态的形式作进一步系统研究。

2. 情态动词的被动语态使用分析

英语议论文中 be 动词被动语态是最为常见、最为基本的被动语态表达方式,除了

be 动词的被动语态之外,情态动词被动语态也是一种使用形式<sup>[17]</sup>。进行英语议论文被动语态研究,情态动词的被动语态研究也是重要部分,为了检验国内英语学习者与母语

者在使用情态动词被动语态时是否呈显著性差异,笔者利用 Antconc 3. 2. 1 软件及 SPSS 的卡方检验来对数据进行进一步统计分析,结果如表 3 所示。

表 3 中国英语学习者和母语者情态动词被动语态的使用对比

情态动词	使用频次		情态动词被动句占总频率百分比/%		显著性检验	显著性值( sig)
	CLEC 语料库	母语者	国内学习者	母语者		
should	97	63	17. 90	17. 50	38. 33	0
can	213	110	39. 30	30. 56	56. 64	0
would	20	39	3. 69	10. 83	11. 88	0. 001
will	105	49	19. 37	13. 61	32. 44	0. 045
could	22	38	4. 06	10. 56	27. 85	0
must	62	36	11. 44	10. 00	12. 84	0. 620
may	18	19	3. 32	5. 28	6. 67	0. 001
might	4	6	0. 74	1. 67	0. 97	0. 988
total	542	360	100. 00	100. 00	187. 62	0

基于表 3 的统计结果进行分析,情态动词 might 和 must 的显著值均大于关键值 0. 05,连接的被动语态形式的使用频次上,国内英语学习者与母语者没有呈显著性差异,表明国内英语学习者能够熟练或擅长运用情态动词 might 和 must 的被动语态表达推断、可能、必须等情态意义。其他的情态动词的显著值均小于关键值 0. 05,均出现显著性差异,这证明国内英语学习者尚未熟练掌握其他情态动词的被动语态。比较特殊的是国内英语学习者过度运用了情态动词 can, will 及 should 引导的被动结构,同时对 might 引导的被动语态又使用得过少。究其原因国内学习者在被动语态的学习中经常将 might 等同于 may 的过去式,从而忽略了这些情态动词还可以表达现在或将来时间的情态含义。例如,学生们经常使用的情态动词 could 不仅是 can 的过去式,还可以用于其他时态,主要是暗示对事情的判断更不确定的意义,除了上述功能之外, could 还可以表示委婉态度等;另外一个情态动词 would 不仅是 will 的过去式,也能表达委婉的语气态度。

根据表 3 的统计分析结果,国内英语学习者对情态动词的被动句的能产性从高到低依次是 can, will, should, must, could, would, may 和 might。然而母语学习者的情态动词被动语态能产性从高到低的顺序依次是

can, should, will, would, could, must, may 和 might。从两个语料库的对比来看,用 can 引导的被动句的使用频次都是最多的,但是相对来说,国内英语学习者的使用频次更高,占总数的 39. 30%。国内英语学习者常用的是 will, should 和 must, 分别占到语料库中的 19. 37%, 17. 90% 和 11. 44%。母语语料库中,位居前三位的是 can, should 和 will, 分别占总语料库的 30. 56%, 17. 50% 和 13. 61%。由此可见,国内英语学习者使用情态动词被动语态时存在过多或过少地倾向于某个情态动词的情况, CLEC 语料库中总的国内英语学习者被动句的使用频次为 542, 母语语料库中使用频次为 360, 国内英语学习者比母语者高出 182, 这种情况显示出国内英语学习者的被动语态使用频次高于母语者被动语态的使用频次。通过数据表明,有超过三分之一的情态动词的被动句被过多地使用。

基于上述分析,个别的情态动词使用是否具有显著性差异已经通过 Antconc. 3. 2. 1w 搜索和检验出来,但在总体样本分布的情况中,国内英语学习者与母语英语学习者在被动语态上是否呈现显著性差异,则需要运用配对样本 T 检验来进行检验,结果如表 4 所示。

通过表 4 的数据显示,均值为 -1. 934 5, 标准差为 8. 069 0, 标准误是 0. 652 1, T 值为 -1. 568, 自由度为 72, 显著值是 0。基于以上



数据显示,因为显著值为0,小于关键值0.05,所以国内英语学习者与母语者在情态动词被动语态的运用上呈现出显著性差异。

表 4 CLEC-LOCNESS 配对样本 T 检验

配对差异检验					T 值	自由度	显著性（双侧）
均数	标准差	标准误	95% 差异置信区间				
			低	高			
-1.934 5	8.069 0	0.652 1	-2.689 7	-0.528 7	-1.568	72	0

根据语法特点,情态动词 should 和 can 常常表示为肯定和坚决的语气,而 would 和 could 表达疑惑和推测的语气,常常用来表明逻辑上存在可能性,而国内英语学习者在情态动词被动语态习得中,往往没有充分考虑到情态动词的语义范畴。因此,国内英语学习者在进行情态动词的被动语态习得时,不仅需要清楚情态动词的语法形式,还要重视情态动词的多层含义,这样才能恰当准确地运用情态动词。

同时,国内英语学习者在英语写作中有通过 should 和 can 等情态动词强加于读者的倾向,具有使读者在阅读中接受自己观点的倾向性。此外,国内英语学习者在英语的学习中普遍最先习得情态动词 can,对 can 的使用更为熟悉,便于在使用中提取,所以国内学习者使用 can 的频次明显高于母语者。国内英语学习者英语议论文中过多使用 should 和 can 的被动结构,而较少使用 would 和 could 的被动结构,这必然会导致学习者议论文中情态动词的表达不足,使读者感觉文章语气生硬甚至几乎带有命令和武断的口吻。

四、结 论

1. 国内英语学习者被动语态使用情况

(1)在 be 动词的被动语态中,国内英语学习者与母语英语学习者的使用情况表现出显著性差异。究其原因国内英语学习者在议论文写作中已经感知到被动语态会使文章更加正式化而不是口语化,因此,以前的议论文写作中被动语态运用不足的状况得到了缓解,但相比之下,众多国内学习者刻意追求运用被动语态使文章更加正式化,造成了国内英语学习者过多运用被动语态的情况。

(2)在情态动词的被动语态中,国内英

语学习者使用典型的情态动词 may, will, might 和 must 的频次和母语者在被动句的使用频次上没有显著性差异,这一点证明了国内英语学习者熟练地应用 will, may, might 和 must 等情态动词来表达将要、可能、必须等情态意义。国内英语学习者与母语者在被动语态使用中的不同则是国内英语学习者使用 should, can 的被动结构过多,同时, would 和 could 的被动结构又使用得过少。

2. 国内英语学习者被动语态问题原因

(1)国内学生在情态动词的被动语态习得中,忽视了情态动词的语义范畴。因此,国内英语学习者在进行情态动词的被动语态习得时,不仅需要清楚情态动词的语法形式,还要重视情态动词的多层含义,这样国内英语学习者才能恰当地运用情态动词的准确意义。

(2)国内英语学习者经常运用 should, can 等情态动词,这些情态动词主要表现为作者将自己的观点强行加给读者,从而让读者不仅易于接受自己的观点,而且具有强迫性,是读者不得不接受自己的意见。

(3)对于国内英语学习者来说,普遍最先习得了情态动词 can 的运用特点,对 can 的应用更有把握,更容易在运用中习得这样的情态动词,所以 can 的使用频次明显高于母语者。其议论文写作中,由 should 和 can 引导的被动句使用过多而由 would 和 could 引导的被动句则使用较少,导致情态表达在议论文写作中明显不足,使文章语气生硬而且几乎带有命令和武断的口吻。

参考文献:

[1] 陈万霞. 从中国学习者英语语料库看英语被动语态习得[J]. 外语教学与研究, 2002(5):

- 198-202.
- [2] 吴颖, 阙紫江. 基于语料库的中国学生议论文写作中的被动语态习得研究[J]. 外语教育, 2013(6): 107-114.
- [3] 刘敬伟, 冯宗祥. 我国英语专业研究生学位论文被动语态的使用分析[J]. 沈阳农业大学学报(社会科学版), 2010(12): 327-329.
- [4] 肖忠华, 戴光荣. 语料库在语言教学中的运用: 中国英语学习者被动句式习得个案研究[J]. 浙江大学学报(人文社会科学版), 2010(7): 189-200.
- [5] 王敏, 李丽霞. FAO 农业英语被动语态的语料库研究[J]. 科技视界, 2013(5): 29-30.
- [6] 马宇立. 中国英语学习者被动语态使用错误分析: 基于 CLEC 语料库的研究[J]. 山东外语教学, 2002(3): 46-48.
- [7] 何安平. 语料库辅助英语教学入门[M]. 北京: 外语教学与研究出版社, 2012.
- [8] 黄昌宁. 语料库语言学[M]. 北京: 商务印书馆, 2007.
- [9] QUIRK R A. Comprehensive grammar of the English language [M]. London and New York: Longman, 1985.
- [10] 章振邦. 新编英语语法教程 [M]. 上海: 上海外语教育出版社, 2000.
- [11] 王克飞. 中国英汉平行语料库的设计与研制[J]. 中国外语, 2012(6): 23-27.
- [12] 宋红波, 王雪利. 近十年国内语料库语言学研究综述[J]. 山东外语教学, 2013(3): 41-47.
- [13] 周世界. 语言统计理论与实践[M]. 大连: 大连海事大学出版社, 2013.
- [14] 戈玲玲, 李广伟, 王一鸣. 基于本源概念: 双语平行语料库的翻译教学平台建设及其教学模式研究[J]. 外语界, 2015(4): 11-17.
- [15] 胡江. 对大学生习得英语现在完成时的语料库研究[J]. 解放军外国语学院学报, 2010(6): 55-60.
- [16] 桂诗春, 冯志伟, 杨惠中, 等. 语料库语言学与 中国外语教学[J]. 现代外语, 2010(4): 419-426.
- [17] 肖中华, 戴光荣. 语料库在语言教学中的运用: 中国英语学习者被动句式习得个案研究[J]. 浙江大学学报(人文社会科学版), 2010(4): 189-200.

## Contrast Analysis of Corpus-Based Chinese English Learners' Passive Voice Usage

LIN Hao<sup>1</sup>, WU Mingfei<sup>2</sup>, FENG Yang<sup>1</sup>

(1. School of Foreign Languages, Shenyang Jianzhu University, Shenyang 110168, China; 2. Department of Basic Subject, Shenyang Urban Construction University, Shenyang 110167, China)

**Abstract:** This paper mainly selects native speakers' corpus LOCNESS as well as non-native speakers' corpus CLEC, applies corpus retrieval software AntConc3. 2. 1w to retrieve and uses corpus analysis software to search, retrieve and further analyze passive voice characteristics in corpus data. Meanwhile, the paper adopts social statistics software SPSS11.5 to analyze and contrast data. It mainly applies quantitative analysis to further contrast and analyze passive voice in English argumentative papers, summarize its existing problems and analyze reasons in order to provide references for domestic English learners' passive voice usage.

**Key words:** passive voice; corpus; second language acquisition; using characteristics