

基于语料库的英语写作任务研究

马凯悦 周丽姣 王思遥

宁夏大学外国语学院 宁夏 银川 750021

【摘要】：英语词汇与写作在英语学习者的学习过程中具有十分重要的地位，基于此，本研究对宁夏大学外国语学院英语专业学生英语写作课堂中为期一个学期的课堂写作任务和课后写作业的词块进行收集和整理的文本 59 篇，约 25000 词建成小型语料库进行调查研究，本文总结基于语料库方法英语写作词块研究的多个方面探讨英语专业的学生与英国本族语者在写作方面的差异，以及语料库在英语写作课程中的应用，并讨论其中的不足以及今后研究可发展的方向，为基于语料库的英语词块研究提供帮助，促进英语写作教学研究的进一步发展。

【关键词】：英语专业；写作教学；小型语料库

1 背景介绍

1.1 英语写作教学的重要性

写作作为英语专业的核心课程之一，是衡量学生英语语言综合能力和体现学生语言应用能力的重要考察方面。因此如何有效地提升学生英语写作水平也是国内外英语教学学者的关注点。传统的英语写作教学是基于老师自己的写作经验对学生的作文进行评价，存在教学方法、教学手段、教学模式单一的问题（乔琳，2018）。随着电子化信息化的发展，为了适应更快发展的信息技术和社会洪流，大学英语写作教学的改革亟待实现。

语料库的载体为电子计算机，承载的是语言知识及素材，基本组成要素为若干的电子文本，它的语言材料具有真实性、代表性，这些独特优点使得其成为专供语言研究的优质资料库（闫亚杰，李欢欢，2015）。将语料库运用到大学英语写作教学中，语料库首先为教师与学生提供了真实的语料；学生可以通过已建的大型语料库获取使用频次最高，表达最为地道的单词、短语搭配等。教师也可以将学生语料与大型语料库进行对比，从而更进一步地了解学生现阶段的写作情况。

1.2 研究目的和意义

本文主要探讨英语专业的学生与英国本族语者在写作方面的差异，以及语料库在英语写作课程中的应用。通过语料库这一创新教学工具的运用，我们希望教师能够有效提高课堂教学效率，为写作教学提供新的思路，激发学生的写作兴趣，引导学生主动的参与写作课堂；学生可以运用语料库加深学习印象，提升写作水平。本研究整理并收集了英语专业学生的写作用本，自建小型语料库。相比大型语料库，小型语料库更加具有针对性，老师可以发现学生普遍存在的写作特征和写作问题。另外，自建语料库可以实时进行语料的

更新，老师也可以对学生的写作能力提升进行更加客观的观察。本文将通过学生写作语料库与成熟的英国本族语者语料库（British Native Corpus）进行对比，从用词准确度、词汇使用情况、词语搭配和句法特征四个方面，进行量化分析，梳理学生写作中的问题，并探讨以上四方面与学生写作质量之间的关系。

2 文献综述

20 世纪 90 年代，Tim Johns 将语料库带入了外语课堂。他认为作为一个大型数据库可以为英文写作提供有效的支撑。作为语料库的奠基人之一，他坚持语料库可以促进以学生为主的学习方式，并由此提出数据驱动学习（Data-Driven Learning）这一概念。学生可以通过积极参与和课堂活动更加有效地学习词汇，提高对词汇的记忆度并将词汇合理地运用在写作中。

国内针对语料库语言学的英文写作研究多基于现有的大型语料库进行分析，且针对特定任务文本的研究。如陈钰婷（2017）以议论文写作为研究对象，分析了议论文的局部衔接，可读性及句法特征。王清华的（2012）将英语专业八级考试中的写作议论文作为分析对象，探讨了不同水平的学习者不同的写作特征。因此，本次研究将研究对象不仅针对议论文，且涵盖了英文写作中其他体裁。

另两篇文献都是基于特定词块和特定对象为研究点，兰婷慧（2012）的研究是基于大型语料库，聚焦固定词块“in my opinion”。叶凯青（2016）研究了学生英语写作中的常见错误。本次研究将不局限与具体词块的研究，而是根据数据分析，找出出现频次较多，具有代表性的词块进行分析。同时，研究对象也是基于宁夏大学的英语专业本科生的习作。闫亚杰，李欢欢（2015）关注语料库的应用，但并未对大学生的英文写作进行具体的数据分析。

综上所述,研究者可以用定量分析、定性讨论的方法,利用大型语料库来进行研究,但同时小型语料库的建设也是十分必要的。因为大型语料库的数据范围太广,缺乏群特征,小型语料库可以服务于特定的使用人群。所以本次研究将根据研究对象的实际情况建设语料库,开展更具有针对性的研究。并且,本次研究不仅仅局限于特点词汇,词块的运用,同时也会根据数据结果整理词频,有针对性地进行研究。

3 实验工具

语料库的构建首先从语料库的设计开始,因为语料类型、大小等都取决于研究目的,并且要确保语料具有代表性,可以反映出学生在写作过程中的问题。因为语料的收集与语料库的结构设计息息相关,为确保收集的语料足具代表性,研究者需要调整语料库的结构。因此在构建语料库前,研究者对所有文本进行随机抽样,主要检查文本的大小,是否有输入错误或者乱码问题等。

后需为方便对语料词汇和句法的分析,研究者需要对所有文本进行词性赋码。词性赋码是指软件根据文本的前后语境,对语料中的每一个单词添加对应的词性,赋码后的文件以 pos 格式保存,因此研究者可以根据词性赋码符号表达句子结构。

经过文本的筛选,本次语料库的构建收集了英语专业的写作共 25817 字,且加入参考语料库英语国家语料库(BNC)作为对比。学习者语料库的写作用本均来自学生的真实写作用本,因此具有较强的真实性和代表性。本次的文本类型则来自于学生的课堂写作用本及课后作业,皆是书写文本。

这些本文可能因手工录入的问题可能会在格式或标点等方面出现错误,因此在导入文本进行分析前首先要对文本进行处理。在本次的研究中,除去人为的文本处理,我们还利用了语料赋码软件 TreeTagger 对语料进行处理。

4 研究过程

4.1 研究问题

本次研究采用量性与质性的研究方法,关注词汇的准确度、词汇搭配情况和句法结构三个方面,通过计算机软件分析自建语料库中的语言现象。并在分析过程中加入英语国家语料库进行对比,阐述其语言现象的异同和原因。本次的研究问题如下:

(1) 自建的学习者语料库与英国国家语料库的高频词汇、短语搭配以及句法结构分别是什么,造成其差异的原因有哪些?

(2) 以上三个因素的使用是否与学生的作文质量有关系?

(3) 通过以上方面的分析,是否会对老师的写作教学和学生的写作水平的提升提供帮助?

4.2 研究步骤

在本次研究中,语料来源是某校英语专业大一学生为期一个学期的课堂写作用本和课后写作用本,共收集文本 59 篇,约 25000 词。

在利用软件 TreeTagger 进行文本的赋码后,笔者将文本导入一款免费的语料库分析软件 AntConc3.5.9。根据研究问题,研究者首先要描述学生用词的准确度。将自建的语料导入 AntConc 后,生成高频词汇词表,并导出数据,之后通过功能检索分析、观察周围词和 N 词序列检索,便可描述词汇运用的准确度。之后利用软件的搭配功能分析某个检索词左右出现的搭配情况,通过对比即可知道学生对于不同词形的使用是否丰富。最后我们利用标注过的文本进行句法的分析,此研究只进行了被动句式和宾语从句的两种分析。

5 研究结果

5.1 用词准确度分析

首先,该研究利用软件的词频统计功能提取了自建语料库中出现的高频词,我们以动词 write 为例,该词在自建语料库中共出现 26 次,之后通过 Concordance 功能查找该词在语境中出现的情况,通过对比即可了解学生对该词的使用情况。

在自建的语料库中,笔者首先观察 write 周边词的前后搭配情况。write 的左边搭配词有介词 to 6 次,人称代词 2 次,also 1 次,and 出现 7 次,can 5 次,should 1 次,then 2 次,that 1 次;write 右边出现的第一个搭配词都是冠词 the, a, an, 出现的第二个搭配词有 outline 7 次, final 8 次, first 6 次, paragraph 1 次, essay 3 次, logical, good, 和 piece 各一次。因此可以总结出学生在使用 write 时,基本都是固定的动宾短语。而在 BNC 中进行检索时,出现在 write 左边的五个高频词分别为 to, I, read, please 和 you; 出现在 write 右边的五个高频词分别为 to, down, about, letter 和 book。由此得出,学生在使用 write 时,鲜少使用介词搭配,多以名词搭配为主。而以英语为母语者的写作中,write 后面多跟介词再接名词。因此在后面的写作课程中,可以更加关注学生对于动词和介词的搭配。

5.2 词语搭配统计分析

表 5.2 自建语料库与 BNC 高频形容词统计表

学生作文自建语料库			British National Corpus		
序数	频数	单词	频数	序数	单词
1	52	beautiful	1	233190	all
2	25	first	2	189133	more
3	17	most	3	144155	some
4	17	some	4	115920	new
5	12	important	5	111578	first

根据 1991 年 Sinclair 对搭配 (Collocation) 一词的定义, 搭配一种短距离内两个或两个以上的词的组合与共现, 它对词义的理解至关重要。观察词汇搭配主要有两种方法, 一是直接观察语境中的关键词 KWIC (Key Word in Context) 检索行; 二是采用统计方法 (Z 值, T 值或 MT 值) 测量节点词与有关词在文本中的共现规律。因此本次研究可以通过 AntConc 中的搭配词功能提取检索词的左右搭配词频数及位置分布, 从而了解搭配词的使用情况。

我们以介词 in 为例, 结果显示与 in 搭配的形符数为 446, 类符数为 1177。出现在 in 左边的搭配词为 believing 共 7 次, located 共 5 次, province 和 park 共 4 次, walk 为 3 次; 出现在 in 右边的搭配词 the 共 64 次, winter, autumn 和 a 各为 3 次。我们可以观察到与 in 的搭配有 believing in, located in, walk in 等, 这些搭配的使用都是正确的。

之后本文分析了形容词的使用搭配情况。自建语料库的高频使用形容词如下表 5.2。我们可以看出一些形容词的两个不同的语料库中都是高频词汇, 因为此类词汇作为限定形容词考察意义不大, 而 beautiful 虽然是使用频次最高的形容词, 但由于该词多出现在命题作文中, 因此极大的影响了该词的使用频率。因此我们选择 important 作为检索词。在自建语料库中该词汇左边的高频搭配词分别为: most, is, an, so, to 等; 出现该词右边的高频搭配词分别为 she, to, role, province, so 等。之后我们采用定量统计方法, 运用 T 值搭配计算, 得出 important 的显著搭配词。学生所使用的搭配词越多, 说明学习者的应用能力越强。因此我们得出结论, 学生可以正确使用 important 的典型搭配用法, the most

important, 但在词汇的搭配丰富性上不足。

表 5.3 important 的显著搭配词

学生作文自建语料库				British National Corpus			
左搭 配词	T 值	右搭 配词	T 值	左搭 配词	T 值	右搭 配词	T 值
most	2.6320 5	she	2.61270	is	82.7 569	to	50.50 99
an	1.3644 5	to	1.57691	an	74.6 973	part	30.98 67
is	1.2196 9	role	0.99787	most	66.7 008	role	24.92 32
so	0.9488 2	provi nce	0.94882	it	53.6 072	than	24.84 04

5.3 句法结构分析

在句法结构的分析中, 我们将通过被动句和宾语从句的观察分析学生的句法使用情况。首先, 我们利用 Treetagger 软件对语料库的文本进行词性标注, 后在 AntConc 中重新导入标注后的文本。根据英文词性对照表, 我们得出被动语态的正则表达式为 “?*_VB*@?*_V?N”, 之后在软件中进行检索, 结果显示被动句 37 条。而宾语从句的正则表达式为 “?*_VV*@that_IN”, 结果显示共 5 条。这两个句型可以揭示学生写作的语言特征, 以及对不同句型的掌握情况。两条数据结果显示, 学生的写作中很少会运用较为复杂的句型, 基本以简短的简单句为主, 这也限制了学生对个人观点的阐述。

综上所述, 我们发现学生的写作在词汇和句法方面仍有所欠缺。在词汇的准确度上, 学生倾向于使用比较简单的词汇。虽然在使用上没有过多的语法错误, 但相较于英语母语者的写作, 学生的表达仍旧不算地道。在 write 一词的使用上, 学生多关注直接宾语的用法, 而鲜少使用介词的搭配。而在用词搭配方面, 学生会较多的使用某种固定搭配, 因此结果显示使用频次高, 而忽略该词的其他表达, 因此在词汇的搭配丰富性上有所欠缺。在句法方面, 学生在写作中句式结构较为简单短小, 不利于自我立场的表达。

6 启示

对于从事二语写作教学的老师, 自建语料库可以有效地通过量性分析发现学生在词汇和句式在使用上的写作特点, 通过与其他大型语料库的对比, 可以更加快速地总结出学生

作文中需要改善的方面,比如老师可以选择在母语者中使用频次最高的表达去替换学生作文中不地道的表达。另外自建语料库虽然过程比较麻烦,且语料的收集需要一定的时间积累,但是也可以提高老师的教学研究意识,从而更好地发现问题解决问题。

学生也是语料库构建过程中重要的一个环节。首先相较于传统的教学方法,语料库可以提升学生写作的自我修改意识。比如老师可以将语料库的分析结果反馈给学生,学生便可以利用语料库对比自己与他人的写作文本找出不足。同时语料库对学生写作表达的积累也非常有必要,随着大型语料

库的更新,学生可以随时查找到最为地道的母语者表达。

本次研究也有许多不足之处。首先是没有收集更多学生的语料文本,语料文本越多,其代表性就越强,并且此次本文的收集类型不够丰富,因此在之后的研究中,可以更加关注学生之后的写作文本,通过纵向对比,也可以发现学生的写作特点和水平提高的因素。其二是语料库的分析结果相较浅薄,没有对句法结构的进行更深入地分析,因此笔者希望可以将语料库应用在之后的文本分析中,从而帮助学生作文水平的提高。

参考文献:

- [1] Guy Aston. Corpora in Language Pedagogy: Matching Theory and Practice, in Principles and Practice in Applied Linguistics. [M]. New York: Oxford University Press,1995.
- [2] Hunston. S. Corpora in Applied Linguistics[M]. Cambridge University Press, 2002.
- [3] Johns, T. Should you be persuaded—two samples of data-driven learning materials[J]. English. Language Research Journal,1991,(4).
- [4] Sinclair. J. Corpus, Concordance, Collocation[M]. Oxford: Oxford University Press,1991: 70.
- [5] 陈钰婷.语料库语言学视角下的中国英语学习者议论文写作多维度分析[D].广西大学,2017.
- [6] 高等学校外语专业教学指导委员会英语组.高等学校英语专业英语教学大纲[M].北京:外语教学与研究出版社,2000:5.
- [8] 兰婷惠.基于语料库的中国英语学习者 in my opinion 语块使用的对比研究[D].西北农林科技大学,2012.
- [9] 乔琳.大学英语专业写作教学模式探究[J].传播力研究, 2008,2(22).
- [10] 王清华.英语专业学生高低分作文中的语言特征分析[D].天津科技大学,2012.
- [11] 叶凯青.基于语料库的大学生英语写作常见错误及成因分析——以浙江工商大学英语本科学生习作实例为例[J].中国教育技术装备,2016(01):70-72.
- [12] 闫亚杰,李欢欢.语料库在大学生英语词汇学习与写作中的应用[J].海外英语,2015(15):93.

作者简介:马凯悦(1998—),女,回族,新疆乌鲁木齐人,文学硕士,单位:宁夏大学外国语学院外国语言学及应用语言学专业,研究方向:英语教学法

通讯作者:吴蓉,女,汉族,教授,单位:宁夏大学外国语学院外国语言学及应用语言学专业,研究方向:英语教学法。