

语料库数据驱动下词块能力与英语写作研究

林巧文, 郑 静

(福建师范大学福清分校 外国语学院, 福建 福清 350300)

[摘要] 词块是英语写作的重要组成部分。基于语料库数据驱动理论, 以非英语专业学习者 of 研究对象, 通过 16 周的实证研究发现, 词块能力与英语写作有显著的相关性; 词块习得不仅能有效地提高学习者的词块能力, 尤其在聚合词和搭配词方面, 而且能提升学习者英语写作用词的多样性、准确性和连贯性。

[关键词] 语料库; 数据驱动; 词块能力; 英语写作

[中图分类号] H 319.13

[文献标识码] A

[文章编号] 1008-889X (2017) 03-113-06

近年来, 语料库数据驱动学习方法是随着语料库语言学的发展逐渐形成的。从本质上来看, 语料库是指通过计算机把大量真实的语言素材, 按一定的语言原则, 采取随机抽样的方式收集起来的自然连续语言^[1]。

数据驱动学习 (Data-driven Learning, 简称 DDL) 是在大量真实语料基础上进行的探索式、自主式的学习模式。它以学习者为中心, 基本理念是鼓励学习者从真实的语言环境中发现语言事实和语言规律, 从而实现自下而上的归纳式学习方法^[2]。从语料库数据驱动的角度来看, 词块可被界定为在某一语域频繁出现的固定或半固定的词组形式, 并传递稳定的语用信息^[3]。语言学家 Sinclair 进一步丰富词块内涵: 结构上固定、语义上完整的词块是语言描述的中心, 并可以作为整体组合被储存、记忆、提取和应用^[4]。

在大学英语教学活动中, 英语写作一直都是教学的薄弱环节。学习者在语言输出中, 不仅要储存大量的词汇、掌握牢固的语法, 而且必须具备一定的英语思辨能力和篇章构建能力。语料库数据驱动的研究发现, 语言交际并非通过单词来实现, 而是依靠那些固定或半固定的“板块构造”^[5]来完成的。因此, 借助语料库, 研究词块能力与英语写作将有益于提高语言学习的效率, 增强语言的地道性、准确性和流利性。

一、国内外研究动态

数据驱动有两种学习模块: (1) 使用“原始语料” (a raw corpus), 师生共同钻研真实的素材, 探索未知结果, 这种研究结果也可能与预设不一致。(2) 教学工作者为某种特定目标而选定索引语料来展现目标语言特征 (the target language feature)。这两种学习模块能培养学习者的学习兴趣, 激发学习者的学习动机, 促进自主学习的形成。正如 Hunston 所指出: “数据驱动”学习不单是教学工作者向学习者具体教授某一语言特征或语言知识, 更重要的是向他们提供语言事实和材料, 要求他们根据这些材料提出假设并最终得出结论。这整个过程都是以“学习者”为中心而设计的一系列教学活动, 目的在于提高学习者对相关知识的敏感度 (consciousness-raising)^[6]。

基于数据驱动的研究, 国内外的专家学者在词块和写作方面做出了坚持不懈的探索和钻研。归根结底, 可以归纳为以下三个方面: (1) 侧重基础研究和理论科研, 以语料库为基础, 阐述词块界定、范畴、维度以及英语写作类型和模式等。基于 Brown Corpus 的研究, Kjellmer 指出语料库能为词块研究提供详细完整的信息资料和数据保

[收稿日期] 2016-03-10

[修回日期] 2016-04-19

[基金项目] 福建省教育厅 A 类人文社会科学研究项目 (JAS14337)

[作者简介] 林巧文 (1979—), 女, 福建古田人, 副教授, 硕士, 主要从事语料库语言学研究。

障,初步提出词块的概念^[7]; McCarthy 进一步明确了词块的概念,将它定义为“习惯性共现的词语”^[8]。而国内学者杨永林在 80 万英语写作语料库上搭建写作平台,倡导“数字化学习理念”^[9]; 滕延江、刘世铸概述了基于语料库英语写作教学新模式^[10]。(2) 探求应用手段和方法,主要研究语料库应用于词块和写作方面的途径,包括词块的检索方式、词块提取、写作阶段、写作方法、科研数据等等。张霞基于中国高级英语学习者研究了词块抽取方案和赋码方案^[11]; 李文中、濮建忠探索了语料库的索引技术^[12]; 陈伟、许之所依托网络语料库,剖析了英语写作的训练方式^[13]。(3) 考察三者结合的实验效果,尝试分析语料库数据驱动的作用、词块能力的形成、英语写作三者相辅相成、相互影响的关系。王立非、张岩通过平行语料库的对比研究发现,在大学英语写作中动词词块和名词词块最为常见,存在过度使用现象^[14]; 在语料库的辅助下,冉玉体的研究也表明词块意识越强,学习者写作能力越高^[15]。

总体上来说,尽管研究取得一定的成效,但仍然存在着一些不足之处:词块能力和写作方面知识性、综述性研究偏多,相关实证研究方面没有取得突破性进展;其次,把语料库数据驱动作为一种新的科研手段和研究方法,并同时考察词块能力和英语写作的相关性研究,研究数据和材料都较少。综合以上因素,本研究拟以语料库“数据驱动”为依据,通过实证研究,结合统计分析软件从定性和定量相结合的角度来探索词块能力和英语写作之间的相关性。

二、实证研究设计

(一) 实验研究内容

本实验是在福建省某高校教育学院 2012 级学前教育专业两个自然班进行的,分别为实验班和控制班,每班人数为 60 人,周期为 16 周。学习者采用的教材为《新视野大学英语》第二版。因此,采用了《新视野大学英语》教材语料库。该语料库信息翔实,语料真实,检索方便,语料覆盖本套系列教材中原文、单词词组、翻译练习、阅读写作及其他辅助性材料。

实验假设为语料库数据驱动的学习环境中,

学习者词块意识增强,词块能力提高,学习者写作成绩也相应提升。回答的具体问题有:(1) 语料库数据驱动的词块能力能否提升学习者写作成绩?(2) 词块能力与写作成绩是否有相关性,如果有,有多大程度上的相关?(3) 如果学习者写作水平确实有所提高,哪些词块类型表现较为突出,哪些词块类型表现较为薄弱?

(二) 实验研究方法

“从科学办法的角度,语料研究办法是一种更强有力的办法,因为其结果是可以验证的”^[16]。因此,本实验的研究方法是:通过实验班和控制班的学习者对词块能力测试和英语写作测试中的成绩差异,结合 SPSS 17.0 的统计数据进行分析,得出科学合理的研究数据。

1. 词块能力测试。目前,二语词块的能力测试主要是围绕目标词而进行的中心词块评估,包括接收性词块和输出性词块,测试词块知识主要通过衡量词块数量和词块知识深度^[17]。本实验主要根据 Michael Lewis 的词块分类而展开讨论。从功能和结构角度上考虑,他把词块分为四种,即:(1) 聚合词 (poly-words)。指传统意义上所说的固定词组,即词块的特殊表现形式,不能随意增减、替换,语义稳定,形式固定,如 in the main。(2) 搭配词 (collocation)。在长期使用中,词与词之间约定俗成的同现关系,受到相应的语法约束,使用频率较高。如副词与动词的过去分词的搭配: well-known, 等。(3) 惯用话语 (institutionalized expression)。指长久以来形成的固定话语组合,包括谚语、名言、警句、社交话语等。如 kill two birds with one stone, 等。(4) 句子框架语和引语 (sentence builders)。指语篇的标识性语言和搭建语篇的句子框架语。如 on the one hand... on the other hand..., 等。词块能力测试卷包括单项选择、词块连线、词块搭配、词块互译;满分为 100 分。本实验在《新视野大学英语》教材语料库中检索出使用频率较高的 500 个词块,按照测试类型组合成两份平行词块能力测试卷。

2. 写作能力测试。英语写作能力测试卷来自历届四六级模拟考题;而评分标准结合英语专业四八级、大学英语四六级和本实验的要求,从内容、结构、衔接、语法、词汇语篇、修辞等方面进行综合评分。为便于比较统计结果,也采取

百分制的形式。为保证评分的公平性和客观性, 每次成绩都请两位资深教师在各自成绩单上填入所得成绩, 再取其平均分。如果写作分数出现3分或3分以上的差别的话, 再由另一位教师一起参与评判。

(三) 实验研究步骤

在实验的16周中, 两个班级同一进度, 同一课时, 保证实验有序开展, 只是在知识传授方面实验班引入了语料库, 采用“数据驱动”的归纳式自主学习方式。词块学习步骤主要包括: (1) 课前词块提取。学习者根据词汇表, 标注出五个左右积极词汇, 通过语料库查询、总结用法。(2) 课堂词块讲授。围绕课文的核心词汇进行讲解并强化训练, 通过语境共现的形式为学习者提供更多搭配和用法。(3) 课后词块巩固。学习者根据所要掌握的词汇进行写作训练、学会融会贯通。如“alter + 名词”的若干词组, 可以查询到以下结果: alter the economic plans, alter one's frugal lifestyle, alter the fact, alter the promotion policy, 等。这种自下而上的归纳式学习词块的方法, 能给学习者迅速搭建一个相互关联的词块表, 在短时间内扩充词块、记忆词块、积累词块。控制班的词块学习不涉及“语料库”方面的知识和应用。课前让学习者预习单词, 熟读并记忆一些核心词汇; 课堂上在语境中讲解新词汇, 培养学习者英语思维的习惯, 给出的范例力求生动形象, 以助记忆; 课后让学习者认真完成词块习题的同时鼓励学习者通过阅读、网络等途径多接触英语词汇, 使学习者处于积极主动的学习状态中。

在写作训练中, 实验班和控制班围绕大学英语每个单元进行单元主题式的写作, 例如, 根据第四册第五单元对照式的写作风格, 让学习者进行“shopping in store and shopping online”写作训练。实验组在语料库数据驱动环境下, 学习者能熟练使用语料库资源, 并通过协作学习探讨交流写作内容, 相互评价写作成果, 形成以教师为辅、学习者为主的写作模式; 而控制组未涉及语料库知识, 依靠传统的结果教学法, 学习者独立完成写作任务后, 提交纸质作文给教师, 教师批改后对于集中出现的问题进行适当的点评。

三、实证研究结果分析

(一) 实验前后英语词块能力和写作水平测试分析

为了科学准确地评判语料库数据驱动的词块能力与英语写作相关性研究结果, 首先要对实验班和控制班进行词块能力和写作水平前测, 看其数据是否满足下一步的实验(见表1)。两个班平均值几乎相同, 实验班的词块能力平均分略高于控制班0.367, 但英语写作方面控制班的平均分(78.361)略高于实验班(77.954)。在方差齐性检验中, 假设方差相同的情况下, 显著性水平设定为0.05, 词块能力测试双尾检测概率Sig. = 0.21 > 0.05, 写作能力测试相应的双尾检测概率为Sig. = 0.30 > 0.05, 这说明两个班级在词块能力测试和英语写作水平测试中, 独立样本T检验均无显著性差异, 符合下一步的实验要求。

表1 实验前两班英语词块能力和写作水平测试独立样本T检验结果

	班级	均值	标准差	均值的标准误	T	Sig. (双侧)
词块能力	实验班	68.324	6.354	1.957	2.369	0.21
	控制班	67.957	6.447	1.654		
写作能力	实验班	77.954	7.214	1.821	2.384	0.30
	控制班	78.361	7.347	1.928		

经过16周测试之后, 再次对实验班和控制班的学习者进行词块能力和英语写作水平的后测(见表2)。从实验数据纵向对比可以看出, 实验

班和控制班在词块能力和写作水平上都有不同程度的提高, 实验班的表现更为突出, 平均分分别高于控制班9.262分和4.119分。设定0.05为

显著性水平,两个班在词块能力和写作水平测试中独立样本 T 检验 Sig. (双侧)均小于 0.05,达到显著性变化。也就是说,在语料库数据驱动

环境下,学习者词块意识得以加强,写作水平有所提高。

表2 实验后两班英语词块能力和写作水平测试独立样本 T 检验结果

	班级	均值	标准差	均值的标准误	T	Sig. (双侧)
词块能力	实验班	78.893	7.240	2.014	2.658	0.00
	控制班	69.631	7.192	1.983		
写作能力	实验班	85.387	7.689	2.235	2.721	0.01
	控制班	81.268	7.820	2.341		

(二) 词块能力对英语写作成绩的影响

为进一步论证英语词块能力与英语写作能力测试之间是否存在相互关联、相互依存的关系,笔者运用 SPSS 17.0 软件对实验班的词块能力后测成绩和英语写作能力后测成绩进行相关分析。

分析结果如表 3 所示,两者的相关系数 $r = 0.618$,显著性水平 0.000,表明达到了统计意义,也就是说,英语词块能力水平与英语写作能力水平之间有一定相关性。

表3 实验组词块能力(后测)与写作能力(后测)的相关分析

		词块成绩	英语写作成绩
词块成绩	Pearson Correlation	1.00	0.618 (* *)
	Sig. (2-tailed)		0.000
	N	60	60
写作成绩	Pearson Correlation	0.618 (* *)	1.000
	Sig. (2-tailed)	0.000	
	N	60	60

从以上数据可得出,英语词块能力与英语写作水平相关。然而,英语词块能力究竟如何影响英语写作测试?笔者运用一元线性回归分析对实验班学习者的后测成绩作进一步分析(见表4)。表4中的“英语词块能力”的 Beta 标准系数为

0.387,它的 T 值达到了显著值,为 0.000 (低于 0.05),说明回归系数是显著的,系数为 0.327。由此可得出回归方程 $Y = 35.954 + 0.327X$,即学习者的英语词块测试每提高 1 分,其英语写作能力测试成绩则相应地提高 0.327 分。

表4 实验组词块能力(后测)和英语写作能力(后测)的一元回归分析

Model	Unstandardized Coefficients		Standardized Coefficient	T	Sig.
	B	Std. Error	Beta		
(Constant) 词块后测	35.954	6.384		4.957	0.000
	0.327	0.957	0.387	7.102	0.000

(三) 实验班和控制班词块使用情况分析

至此,笔者已经回答了前面两个假设,为了回答第三个问题,即:如果学习者写作成绩有所

提高,哪些词块类型表现较为突出,哪些词块类型表现较为薄弱?在写作能力后测中,集中收集学习者的词块运用情况,对学习者使用的聚合词、

搭配词、惯用话语和句子框架语进行统计分析, 得出实验结果(见表5)。实验结果表明, 实验班和控制班在这四类词块运用方面极不平衡。

实验班的学习者在聚合词、搭配词和惯用话语方面均超过控制班, 但句子框架语方面还略低于控制班。不论是实验班还是控制班, 词块运用

方面表现明显的都是聚合词和搭配词, 分别占了各自班级总词块数量的78.2%和69%; 惯用搭配词块的运用方面表现最弱, 总词块数分别为46和38。也就是说, 在英语写作中, 实验班学习者运用词块数量总体上大大超过控制班学习者, 尤其体现在聚合词和搭配词方面。

表5 实验班和控制班使用词块的种类对比

类别	实验班		控制班	
	数量	占比 (%)	数量	占比 (%)
聚合词	274	42.0	110	25.2
搭配词	236	36.2	192	43.8
惯用话语	46	7.1	38	8.6
句子框架语	96	14.7	98	22.4

(四) 实验结果定性分析

研究表明, 词块能力与写作水平密切相关, 高分者具有更加丰富的词块知识, 这个发现支持了丁言仁、戚焱^[18]的观点, 即词块能力与英语写作有显著的相关性。虽然实验班和控制班的学习者在词块运用和写作训练中呈现出相似之处: 在写作中都能切入主题, 把握关键词汇, 熟悉文章结构, 因此在句子框架语方面差异不大, 但两个班级学习者在其他方面还是有较大的差别。

1. 实验班学习者在写作中使用词块数量总体上比控制班要高得多, 如学习者会用 *complete* 代替 *include*; *be scheduled to do* 替代 *arrange*, 等。控制班学习者偏爱用一些简单的单词来表达, 而忽视了许多有意义的词组搭配。

2. 实验班学习者在写作中体现了词块多样性和丰富性, 能够灵活运用像 *where necessary*, *do one's utmost*, *go against* 等词块。而控制班的学习者在写作中几乎没有涉及这些词块的使用。

3. 实验班学习者在写作中词块的选择更加书面化, 而控制班学习者更倾向于口语化表达^[15]。如在表示自己的观点时, 实验班的学习者会使用 *to be of the opinion that*, *hold the opinion*, *as far as I'm concerned* 等; 而控制班的学习者较多使用了 *in my opinion*, *I think*, *I guess* 等表达方式。

4. 实验班学习者在写作中衔接手段更自然、流畅。词块是语篇的重要组成部分, 是语篇衔接

的主要标识。大多数的聚合词和搭配词都具有篇章组织功能, 如转折、举例、总结、对照等, 词块具有完善篇章结构, 提示篇章信息, 暗示篇章内容的功能^[19]。实验班学习者较多运用这种词块, 使文章衔接自然, 浑然一体。如实验班学习者在写作中用到的衔接词块有: *by contrast*, *for instance*, *as a consequence*, *at the same time* 等; 而控制班学习者在词块选择上比较贫乏, 常常使用 *in a word*, *on the other hand*, *what's more* 等。

这些差异的产生可能是由于通过语料库数据驱动学习模式, 实验班学习者在词块理解和运用能力上要高于控制班学习者, 并且学会辨别书面语和口语使用情况和功能特征, 在写作中灵活地贯穿各种词块, 提高英语写作质量。而控制班学习者虽然能够输入词块信息, 但由于缺乏真实的语言环境和丰富的语料资源, 词块的理解只限于机械记忆, 主动认知和运用能力明显不足。

四、结 语

总之, 语料库数据驱动的学习方式能为学习者提供丰富真实的语言素材, 促进词块的理解、记忆、内化与运用, 增强语言输出的准确性和流利性, 切实提高英语写作水平。同时, 利用语料库数据驱动的词块能力能从技术层面和教学理念上弥补以往的英语写作缺陷, 它不仅培养学习者自主学习能力, 而且从根本上改变教学方法。因

此,随着语料库语言学不断深入发展,将词块能力应用于英语写作必将成为未来发展前景之一。本研究还存在一些不足之处:对学习者来说,通过语料库的数据驱动来分析词块搭配、意义,不仅耗时耗力、枯燥乏味,而且面对浩瀚的语料资源,学习者往往无处下手;对教学工作者来说,语料库技术检索和分析能力是一种艰难的挑战,这种研究成果难以转化为教学实践。另外,本试验周期较短,人数较少,区域专业受限,更大规模的实践势在必行。

[参考文献]

- [1] 李文中. 语料库、学习者语料库与外语教学 [J]. 外语界, 1999 (1): 51 - 54.
- [2] 黄大网, 南佐民, 杨新亮. 数据驱动学习与英语写作教学中的语域知识传授——基于 CLEC 语料库的个案研究 [J]. 外语电化教学, 2003 (6): 37 - 41.
- [3] LEWIS, MICHAEL. The lexical approach [M]. London: Language Teaching Publications, 1993.
- [4] SINCLAIR, JOHN. Corpus, concordance, collocation [M]. Oxford: Oxford University Press, 1991.
- [5] 杨玉晨. 英语词汇的“板块”性及其对英语教学的启示 [J]. 外语界, 1999 (3): 24 - 26.
- [6] HUNSTON S. Corpora in applied linguistics [M]. Cambridge: Cambridge University Press, 2002.
- [7] KJELLMER G. Aspects of English collocations [C] // WEISS, eds. Corpus Linguistics and Be-

yond. Amsterdam: Rodopi, 1987.

- [8] MCCARTHY M. Vocabulary [M]. Cambridge: Cambridge University Press, 1990.
- [9] 杨永林. 一种基于数字化教学理念的写作训练系统 [J]. 外语电化教学, 2004 (4): 9 - 14.
- [10] 滕廷江, 刘世铸. 基于语料库的英语写作教学新模式 [J]. 外语电化教学, 2006 (6): 49 - 54.
- [11] 张霞. 基于语料库的中国高级英语学习者词块使用研究 [J]. 外语界, 2010, 4 (5): 116 - 118.
- [12] 李文中, 濮建忠. 语料库索引在外语教学中的应用 [J]. 解放军外国语学院学报, 2001 (3): 20 - 25.
- [13] 陈伟, 许之所. 基于网络资源的大学生英语写作能力培养模式研究 [J]. 外语界, 2008 (4): 34 - 40.
- [14] 王立非, 张岩. 大学生英语议论文中高频动词使用的语料库研究 [J]. 外语与外语教学, 2007 (2): 30 - 34.
- [15] 冉玉体. 语料库辅助下的词块习得与写作能力的相关性研究 [J]. 英语研究, 2013 (2): 68 - 72.
- [16] LEECH, G. The State of the Art in Corpus Linguistics [M] // AIJMER K, ALTENBERG B. English corpus linguistics. London: Longman, 1991.
- [17] 梁岚. 基于笔语语料库的大学英语学生词汇发展研究 [J]. 外语学刊, 2003 (2): 126 - 130.
- [18] 丁言仁, 戚焱. 词块运用与英语口语和写作水平的相关性研究 [J]. 解放军外国语学院学报, 2005 (3): 49 - 53.
- [19] 谢爱红. 词块使用水平与英语写作成绩相关性研究 [J]. 湖南农业大学学报, 2009 (6): 71 - 74.

Correlation Research between Lexical Chunks Competence and English Writing by Data-driven Corpus

LIN Qiao-wen, ZHENG Jing

(Foreign Language College, Fuqing Branch of Fujian Normal University, Fuqing 350300, China)

Abstract: Lexical chunks are one of the important components of the English writing. Based on the corpus - data - driven theory, the empirical study lasted sixteen weeks and took the non - English majors as research objects, which has revealed there is significant correlation between lexical chunks competence and English writing. Not only can it effectively improve the learner's chunks competence, especially in poly - words and collocation, but it can enhance the diversity, accuracy and coherence of lexical chunks in English writing.

Key words: corpus; data - driven; lexical chunks competence; English writing

(责任编辑 陈蒙腰)

