

数据驱动学习视角下的 英语词汇教学探究

徐亮^{1,2}, 张勇¹

(1. 江西农业大学 外国语学院, 江西 南昌 330045; 2. 西安交通大学 外国语学院, 陕西 西安 710049)

摘要:基于语料库的数据驱动学习(DDL)是近年来外语教学研究的热点和重点之一。数据驱动学习以真实语料为教与学的材料,体现“以教师为主导、以学生为主体”的教学理念,注重培养探究型学习意识。以英语中的一对近义词“edition”和“version”为例,探讨数据驱动视角下的语境共现、N元分析、搭配与类联接分析以及网络语料库等几个方面的英语词汇教学方法及应用。结果表明,利用基于网络在线的BNC、COCA以及iWeb Corpus等英语本族语语料库,教师可自建用于教学目的的小型语料库,能够进行有效的近义词辨析等方面的英语词汇教学。

关键词:数据驱动学习;语料库;英语词汇教学

中图分类号:H3

文献标志码:A

文章编号:1674-5884(2019)05-0142-06

早在20世纪70年代,英国语言学家夸克(R. Quirk)、利奇(G. Leech)等就开始探讨语料库语言教学实践,利用语料库资源制定教学大纲、编写教材、进行语言学习等^[1]。1980年,英国学者辛克莱(John Sinclair)等人以COBUILD语料库为基础进行了大量的词汇语法研究,并开发了系列的工具书和教材^{[2][103]}。1991年,约翰斯(Tim Johns)提出了数据驱动学习(Data-Driven Learning,即DDL)的概念,其主要思想是让学生基于大量的语料库数据观察、概括和归纳语言使用现象,是一种基于语料库数据的语言学习法^{[3][4]}。

语料库从诞生之日起就与语言研究以及语言教学研究息息相关。数据驱动学习这一术语就是从语料库语言教学中产生的,它指的是教师和学生语言教与学的过程中,以真实语料为基础,以索引工具软件为手段,对检索得到的真实语料的语义表达、语用特征以及语法规则等语言现象进行观察、分析,目的是促进语言学习^[5]。

数据驱动学习是一种以词为基础的语言学习

方法,随着语料库技术的发展,以词为中心的语境共现、N元分析、搭配与类联接分析及网络语料库等方面的技术、工具软件以及教学理念等日趋成熟。数据驱动学习理念下的外语词汇学习研究也越来越多,我国学者已经做了很多基于语料库的英语学习研究,他们的重点也大多放在英语词汇教学方面。语料库是一个按照一定的采样标准采集而来的、能够代表一种语言或者语言的一种变体或文类的电子文本集^{[2][3]}。语料库为语言教学提供了大量真实、丰富、生动的语言材料。我国外语界的学者(如桂诗春、杨惠中、冯志伟、何安平、卫乃兴、梁茂成等)也积极倡导将语料库应用于我国外语教学的教材编写、语言学习材料技术处理等各个方面,特别是教学加工^[6]。

Boulton进行了一项基于数据驱动学习模式的语法教学研究,被试为132名较低英语水平的法国大一学生,研究结果表明,DDL比传统的模式更加有效,当然这些仍需更多的教学实验进行验证^[7]。而蒋平等人进行的语料库辅助教学实

收稿日期:20190405

基金项目:江西农业大学教改课题(2010B2ZC16;2016B2ZZ29)

作者简介:徐亮(1978-),男,陕西榆林人,讲师,博士生,主要从事二语习得、语料库语言学以及外语教育技术等研究。

证研究表明,语料库的使用在词汇语法教学方面较为有效,DDL对于外语水平较差的学习者有一定难度^[8]。以往研究涉及DDL的特征分析以及网络语料库在DDL中应用的并不多见,本文将探讨数据驱动学习的基本特征,重点阐述数据驱动学习视角下英语词汇教学的应用与实践。

1 数据驱动学习的特征

数据驱动的学习以计算机文本为媒介,采用频数、概率等统计手段,对显性的语言行为进行观察,用归纳、探索等方法学习语言。数据驱动学习模式不同于传统外语教学模式,具有以下几个特征:

第一,以真实语料为教与学的材料。语料库为语言学习的输入提供真实的语言材料,有助于建构真实的语言学习环境。以经验主义为哲学基础的语料库语言学以描写语言事实及语言使用为目标,分析从语境中获取的真实语料。

第二,体现“以教师为主导、以学生为主体”的教学理念。数据驱动的语言学习虽然强调学习者在学习过程中的自我发现、自我探索,但并不是一般意义上的自主学习。教师在这个过程中主导作用也同样不可忽视。不同于西方“以学生为中心”的理念,我国教育学界提出的“教师主导、学生主体”的教育理念认为教学的方向、内容、方法、进程、结果和质量等都主要由教师决定和负责,学习是学生自己的独立的、主动的活动^[9],教学中教师的主导作用和学生的主体地位同样重要。“教师主导”是教师主体服务性的表现,是学生发展的外因,“学生主体”则是在教师主导下发展着的主体^[10]。约翰斯认为,在数据驱动学习中,教师是学生自发研究型学习活动的指导者和协调者^[5]。数据驱动学习中,教师是整个教学过程中方向的把握者、内容的设计者、方法的演示者以及结果和质量的评价者,充分体现着教师的主导作用。学生的主体地位在数据驱动学习中始终未变,他们在教师的引领下进行探索性学习。

第三,培养探究型学习意识。数据驱动学习环境下,学生在教师的组织、引导及示范作用下通过对真实语言使用情况进行观察、归纳、分析以及

整理,是自我探索发现、自我管理的一种自主学习。强调探索和发现的学习过程,主张自下而上、归纳式的学习。数据驱动的外语学习重视学生的学习过程,语料库可提供大量真实的语言数据,学生可获得地道自然的语言输入。教科书中的一些语言使用描述往往并不与实际使用情况一致,数据驱动学习更有利于学生发现这些现象,并掌握真实的语言使用特点。例如,何安平通过语料库检索发现,在JLESSON语料库^①初中英语教材第一册的69例“any”中,44例为否定陈述句,25例为疑问句,无一例肯定陈述句,这与教科书上强调的“any”一般用于否定句和疑问句的说法并不一致^{[11]49-50}。国外学者经过对大量真实语料的调查后发现,有50%的“any”竟是用于肯定陈述句。

2 数据驱动学习与英语词汇教学:方法与应用

数据驱动学习这一概念起源于语料库语言学,是语料库与语言教学结合发展到一定阶段的产物。随着时代的发展,它的内涵和外延进一步扩大,从最初的百万词级语料库扩展到亿词级语料库,从单机版语料库扩展到网络版语料库,从语言研究扩展到语言教与学以及语言教学研究,从词频分析扩展到搭配、意义以及语境分析等。

语音、词汇、句法是语言的三大要素,词汇在语言学习中的作用不言而喻。学习者词汇掌握程度直接影响着其听、说、读、写等各项语言技能的提高,词汇教学在培养学生英语语言能力方面的作用也是显而易见的。如何进行有效的英语词汇教学一直是我国外语教学界关注的一个重要内容。Hadley、Mizumoto & Chujo以数据驱动学习为方法的教学实验表明,这种方法对日本大学生英语语法学习是有效的^[12-13]。Luo、Luo & Zhou的实证研究表明,数据驱动学习在中国大学生英语写作教学中可以有效提高学生语言使用的准确性和流利性、语言产出运用以及促进学生自主学习能力、增强语言使用意识等等^{[14][15]}。辛克莱等学者基于语料库的研究提出了词汇教学大纲,强调英语学习者应学习语言中最常见的词语形式、词语用法的最核心形式以及典型组合

① JLESSON语料库是“中学英语教育教学语料库”(CEEC)的子语料库中的一个部分。CEEC由华南师范大学外国语言文化学院何安平教授及其团队建立,包括两个子库:中学英语教材语料库和中学课堂话语语料库。

形式^{[16]235-237}。

2.1 语境共现

语境共现 (concordance), 也称“语境中的关键词” (key word in context, 简称 KWIC) 或“索引”, 指的是运用索引软件在语料库中查询某词或短语的使用实例, 然后将所有符合条件的语言使用实例及其语境以清单的形式列出^{[2]11}。例如, 卫乃兴、王海华和王同顺通过对“cause”一词的检索并观察索引行及其语义韵 (semantic prosody) 后发现, 该词后接的宾语几乎都是“problems”“damage”“pain”等一些不好的结果^[17-18],

但很多词典并未提及该词的这种意义^{[2]16}。

语境共现还可用于辨析近义词、同形异义词等。例如, 英语中“edition”和“version”都有“版本”的意思, 二者的意义并不好区别。笔者借助语料库, 在“英国国家语料库” (British National Corpus, 缩写为 BNC) 检索这两个词^①并观察其语境后发现, “edition”一般指书籍、杂志报纸等的版本或者广播、电视节目的一期或一辑; 而“version”一般指某一件事情或事物的不同说法、个人的描述等, 或指某种软件的版本。详见图 1、图 2 以及表 1、表 2。

1.	edition_BNC	achievements of women writers, and Colman and Thornton published the first	edition	of Poems by Eminent Ladies. In the 1750s, the place
2.	edition_BNC	you look at scene, act four, scene five in my	edition	. Now I discovered this morning there is actually act four scene
3.	edition_BNC	actually act four scene three in some was it just a-- an	edition	of that (SP: KPVPUSUNK) ? Either a rare or
4.	edition_BNC	that (SP: KPVPUSUNK) ? Either a rare or first	edition	, something like that. ? (SP: KPVPUSUNK) ?
5.	edition_BNC	with that breathless and extravagant interest which demands a copy of every	edition	of the Star'. In Birmingham' the evening papers are
6.	edition_BNC	and Britain's best pub guide. This year's	edition	, once again generously sponsored by British Coal, is no exception
7.	edition_BNC	established a great new vogue in 1836-37. The description 'first	edition	in book form' normally means that the work was first published
8.	edition_BNC	case. The 1693 quarto is loosely described as the 'second	edition	', but strictly speaking it is a reissue of Q1692:
9.	edition_BNC	7. 30pm. # MAKE SURE THIS CAP GOES IN NDOW	edition	. THERE'S ANOTHER VERSION FOR SERIES Jean Walsh (left
10.	edition_BNC	Form of Collateral warranty for purchasers and tenants and in the third	edition	of the form for funders. It commented favourably on a model
11.	edition_BNC	Hornes Collective catalogue makes it all so easy. Our handsome first	edition	covers everything from suits to casual wear. All stylish, all
12.	edition_BNC	CATHEDRAL CHOIR/DARLINGTON # ? (----) ? David Skinner's	edition	of the Threepart Mass employs an upward transposition from the written original
13.	edition_BNC	1983, and a third term in 1987. In the first	edition	of this book it was argued that levels of controversy over social
14.	edition_BNC	as possible, Chapman persuaded the Examiner to publish a Saturday sports	edition	, the first issue of which appeared on 10 January, 1925
15.	edition_BNC	Wales # I've been giving some thought to the next	edition	of our magazine Rural Wales /Cymru Wledig, and I would like
16.	edition_BNC	licensing its overseas Branch offices or other third party to produce an	edition	in return for a royalty payment to overcome special local circumstances the
17.	edition_BNC	Happy return for File # THE Belfast Telegraph's new-look Saturday	edition	with two bumper magazine sections has brought about some changes to your
18.	edition_BNC	best selling dictionary for learners of English. The recently published fourth	edition	was updated with reference to the database of the incomparable Oxford English
19.	edition_BNC	but it certainly does not go far enough; perhaps the second	edition	will improve on this sad state of affairs! # Greed in
20.	edition_BNC	second-hand books' emporium. Southern readers might do better with the	edition	covering their own counties. Anyone taking a detecting holiday in Northumberland

图 1 BNC “edition”索引行 (前 20 行)

1.	version_BNC	correspond to either	version	, (45) can only match the "unacquainted" version
2.	version_BNC	version, (45) can only match the "unacquainted"	version	. (44) how did Oliver imagine her? (45
3.	version_BNC	given to timeshare customers, is 102 words long. His	version	uses words such as buyer and seller instead of offeror and offeree
4.	version_BNC	before, Murder in the Cathedral had been released in its cinematic	version	. He told the American writer, John Malcolm Brinnin, that
5.	version_BNC	prospects. He made a living by journalism, in the 1960s	version	of Grub Street eventually landing the job of drama critic for Town
6.	version_BNC	and the private, often anguished individual. Remainsil became the staged	version	of her perfected life, the culmination of the myth which was
7.	version_BNC	at the Sybase User Conference, Microsoft showed off a prototype 32-bit	version	of SQL Server designed to exploit NT's multiprocessing architecture for
8.	version_BNC	if it is to sing the normal song rather than the crude	version	. (c) At the bottom is the song of a
9.	version_BNC	RATES 66MHz PENTIUM AT 112 MIPS # Intel Corp rates the 66MHz	version	of the Pentium chip at 112 MIPS -- well it is claimed
10.	version_BNC	title song is er? (unclear)? and the second	version	which is a remix version which I've done, which
11.	version_BNC	(unclear)? and the second version which is a remix	version	which I've done, which I was working on last
12.	version_BNC	the game is aimed at older children and adults. A junior	version	of the game is being developed. The game can be played
13.	version_BNC	expanded for the 1747 event. It looks as if the manuscript	version	(with a different ordering from the printed one) was the
14.	version_BNC	to 'screaming' by Wayne (RH). The final	version	, which met with the approval of all pupils, appeared in
15.	version_BNC	its allies in the upcoming Windows NT blitzkrieg, that the server	version	of Windows NT will start shipping by the third week of June
16.	version_BNC	functional form: ? (deleted: formula)? A simplified	version	of this function can be obtained by using real national income (
17.	version_BNC	and we're still rushed off our feet. The Dutch	version	is selling well and we expect the German version to do even
18.	version_BNC	. The Dutch version is selling well and we expect the German	version	to do even better. 'now awaiting my registered version.
19.	version_BNC	German version to do even better. 'now awaiting my registered	version	. Registered users will receive an update to the latest version and
20.	version_BNC	registered version. Registered users will receive an update to the latest	version	and a utilities pack which includes data import/export, file rebuilding and

图 2 BNC “version”索引行 (前 20 行)

2.2 N 元分析

N 元 (N-gram) 是计算机术语, 指文本中连续 n 个词 (或其他单位) 所构成的序列, n 最小值是 1, 也就是一个词 (unigram), n 也可以是 2、3、4、5 等, 对应的 n -gram 分别是 bigram、trigram、4-

gram、5-gram 等, 语料库语言研究中经常使用 N 元, 称为词束或词块 (lexical bundles) 或者词簇 (word cluster)^[19]。例如, 梁茂成经过统计发现, BNC 中最高频的五元组前 6 位的词束及其频率分别是 at the end of the (4 531 次)、by the end of

① 笔者以“edition”和“version”为例, 利用 BNC 语料库中这两个词的检索结果 (edition 2 439 例, version 7 795 例) 建了小型的 BNC-edition & version 语料库。以 BFSU PowerConc 1.0 为检索工具, 单词数定义为 [a-zA-Z0-9-]+, BNC-edition 的字数为 66 599, BNC-version 的字数为 215 797, BNC-edition & version 的总字数为 282 396。本文“3.1 语境共现”“3.2 N 元分析”及“3.3 搭配与类联接分析”等部分的图表内容都是基于自建的 BNC-edition & version 语料库。限于篇幅, 索引行、四元分析等仅列出了排在前 20 行 (个) 的例子。

the(1 840 次)、I don't want to(1 761 次)、I don't know what(1 751 次)、as a result of the(1 597 次)、in the middle of the(1 516 次)。这些高频 N 元组恰恰是语言中最重要的组块,本族语者和高级二语学习者都能够流利准确地以短语单位对复杂语言形式进行存储和加工,这体现了辛克莱的习语原则。我们可以用 N 元分析对教材进行分

析,适当利用词块方法教学。观察“BNC-edition & version 语料库”中的四元分析,就可以归纳出最常用的词块表达,如“in the first/second/new edition of”“the latest edition of (this) (book)”“a/the (new/latest/first/last) version of the (module)”等(见表 1 和表 2)。

表 1 BNC-edition 四元分析(前 20 个)

序号	四元组	频数
1	the first edition of	96
2	the second edition of	51
3	a new edition of	45
4	in the first edition	35
5	of the first edition	35
6	first edition of this	34
7	The latest edition of	32
8	edition of this book	29
9	the new edition of	29
10	's edition of the	27
11	in the second edition	25
12	new edition of The	23
13	the third edition of	23
14	first edition of the	19
15	an edition of the	17
16	latest edition of the	15
17	first edition of his	14
18	second edition of the	14
19	Since the first edition	14
20	to the first edition	13

表 2 BNC-version 四元分析(前 20 个)

序号	四元组	频数
1	version of the module	149
2	a version of the	120
3	a new version of	110
4	the latest version of	81
5	The first version of	74
6	's version of the	68
7	the next approved version	59
8	new version of the	56
9	module version which does	55
10	version which does exist	55
11	first version of the	54
12	the version of the	47
13	the last version of	46
14	does exist in the	45
15	exist in the format	45
16	which does exist in	45
17	should specify either a	43
18	You should specify either	43
19	latest version of the	41
20	last version of the	39

2.3 搭配与类联接分析

早在 20 世纪 50 年代,英国著名语言学家弗斯(J.R. Firth)就提出了搭配(collocation)的概念,指的就是词语之间的“结伴关系”。在语料库索引行中,以检索词为中心左右之和为跨距(span),跨距内每个位置出现的词即节点词的搭配词(collocates)^{[2]12}。与搭配相近的一个概念就是类联接(colligation),它所关注的“结伴关系”不再限于词汇层面,而上升到即语法互选,如形容词经常与名词构成类联接。

笔者利用由许家金和贾云龙开发的 BFSU Collocator 1.0 搭配强度计算工具软件对自建的“BNC-edition & version 语料库”进行了计算,分别以“edition”和“version”为节点词,设跨距为 4,以 Log-log 值对它们的搭配词从高到低进行了排

序,发现“edition”与序数词搭配最多,表示书籍、报纸或电视、广播节目等的“版本”“一期”或“一辑”等;而“version”则不同,与它搭配的一般都是形容词,表示对某种事情的赞同、许可或授权等的描述。这与之前的索引行观察、四元分析等结果是一致的(详见表 3 和表 4)。

2.4 网络语料库

计算机网络技术的发展推进了语料库的发展,“网络语料库”变得越来越普遍。梁茂成认为,网络语料库大致分为两种^{[19]17},第一种是把整个网络上的所有文本或部分文本视作语料库,即 Web as Corpus(WaC),网络作为语料库;第二种是把现成的语料库上传到网络上供用户使用,即 Corpus on the Web(CoW),网络在线语料库,如 BNC、COCA 等。熊文新提出,把 Web 当作语料库

时应考虑对其采集的文本做进一步加工处理,以

构造优化语料库资源^[20]。

表 3 “edition”的搭配词(前 20 个)

序号	搭配词	频数	Log-log
1	limited	169	16.093 9
2	second	388	14.716 5
3	first	952	13.510 1
4	2nd	34	11.378 6
5	Special	130	11.307 9
6	third	159	10.934 3
7	fifth	30	10.320 7
8	paperback	38	10.020 0
9	new	1 073	9.965 2
10	4th	24	9.856 5
11	critical	51	9.726 2
12	revised	201	9.661 5
13	3rd	23	9.424 2
14	fourth	53	9.233 3
15	6th	15	8.991 9
16	5th	17	8.781 6
17	sixth	24	8.780 2
18	29th	13	8.128 4
19	Personal	116	8.095 6
20	morning	46	7.941 8

表 4 “version”的搭配词(前 20 个)

序号	搭配词	频数	Log-log
1	of	12 422	10.318 7
2	final	156	9.774 4
3	approved	111	9.242 5
4	latest	361	9.159 0
5	simplified	49	9.042 6
6	events	121	8.904 8
7	modified	62	7.983 8
8	Japanese	71	7.944 6
9	own	262	7.944 1
10	Authorised	41	7.539 6
11	preferred	50	7.314 3
12	authorized	23	7.108 5
13	updated	91	6.913 9
14	6	156	6.437 4
15	66MHz	18	6.400 4
16	shorter	32	6.296 5
17	cut-down	17	6.231 6
18	shortened	13	6.212 2
19	tested	26	6.197 4
20	film	98	6.102 1

发布于 2018 年 5 月的 iWeb 语料库(The iWeb Corpus)^①是由美国杨百翰大学 Mark Davies 教授创建的迄今为止规模最大、功能强大的英语本族语网络语料库之一,“对于语言学习者、教师和研究者来说,堪称绝佳理想的语料库”^[21],特别适合数据驱动学习。笔者利用 iWeb 语料库自建

了专用的“edition-version”虚拟语料库(见图 3),不但可以进行常规的语料库检索,还可对检索结果进行跨库比较;可以全方位地对词汇的常见搭配、词块、索引行、语境等语言特征进行细致、深入的多模态式观察、分析和探究。

iWeb: The 14 Billion Word Web Corpus						
SEARCH		VIRTUAL CORPORA		CONTEXT		HELP
MY VIRTUAL CORPORA						
HELP	↑	↓	LIST NAME	# WEBSITES	# WORDS	FIND KEYWORDS * SPECIFIC * FREQ
1			EDITION_VIRTUAL	19	3,149,903	NOUN VERB ADJ ADV N+N ADJ+N
2			VERSION_VIRTUAL	19	1,610,317	NOUN VERB ADJ ADV N+N ADJ+N

图 3 基于 iWeb 语料库的“edition-version”虚拟语料库

3 结语

以语料库为基础的数据驱动学习可以让我们通过语境共现、N 元分析、搭配与类联接分析观察真实语言现象、归纳语言使用,揭示某些词语的语

义韵和语义倾向,是信息化时代语言学习的重要形式和途径。当前的网络语料库及其分析工具的强大功能使得英语词汇教学更加科学化和信息化,数据驱动的语言学习也将成为外语学习的一种重要途径。

① iWeb 语料库的库容为 140 亿词,是 COCA(当代美国英语语料库,5.6 亿词)的 25 倍,BNC(英国国家语料库,1 亿词)的 140 倍;仅需数秒,就可建成一个基于任何主题的“虚拟语料库”,然后就可以像单机版语料库一样进行检索、生成词表等;它还兼容各式各样的检索方案,无须专门学习复杂的检索语法,仅需数秒就可进行各种宽泛检索、生成词块序列等;此外,iWeb 语料库还提供了前 60 000 个高频词的发音检索、词典页面(含相关话题的定义、图片、发音、视频等多模态语料)。

参考文献:

- [1] 何安平. 语料库与外语教学[J]. 国外外语教学, 2001(3):15-19.
- [2] 梁茂成, 李文中, 许家金. 语料库应用教程[M]. 北京: 外语教学与研究出版社, 2010.
- [3] 甄凤超. 语料库数据驱动的外语学习: 思想、方法和技术[J]. 外语界, 2005(4):19-27.
- [4] Johns T. Should You be Persuaded—Two Samples of Data-Driven Learning Materials[J]. In Tim John & Philip King, (eds) "Classroom Concordancing", ELR Journal, 1991(4):1-16.
- [5] Johns T King P. (eds). Classroom Concordancing[J]. English Language Research Journal, 1991(4):27-45.
- [6] 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.
- [7] Boulton A. Testing the Limits of Data-Driven Learning: Language Proficiency and Training[J]. ReCALL, 2009(1):37-54.
- [8] 蒋平, 孙晓燕, 徐雪梅. EFL 环境下语料库辅助基础英语教学实证研究[J]. 教育学术月刊, 2013(12):81-86.
- [9] 王策三. 论教师的主导作用和学生的主体地位[J]. 北京师范大学学报(社会科学版), 1983(6):70-76.
- [10] 刘继萍. "教师主导 学生主体": 一个经典公式正当性的论证[J]. 北京师范大学学报(社会科学版), 2015(4):25-33.
- [11] 何安平. 语料库语言学与英语教学[M]. 北京: 外语教学与研究出版社, 2004.
- [12] Hadley G. Sensing the Winds of Change: An Introduction to Data-Driven Learning[J]. RELC Journal, 2002(2):99-124.
- [13] Mizumoto A, Chujo K. Who is Data-driven Learning for? Challenging the Monolithic View of its Relationship with Learning Styles[J]. System, 2016(61):55-64.
- [14] Luo Q. The Effects of Data-Driven Learning Activities on EFL Learners' Writing[J]. SpringerPlus, 2016(5):1-13.
- [15] Luo Q, Zhou J. Data-Driving Learning in Second Language Writing Class: A Survey of Empirical Studies[J]. International Journal of Emerging Technologies in Learning, 2017(3):182-196.
- [16] 桂诗春. 多视角下的英语词汇教学[M]. 上海: 上海外语教育出版社, 2013.
- [17] 卫乃兴. 语义韵研究的一般方法[J]. 外语教学与研究, 2002(4):300-307.
- [18] 王海华, 王同顺. CAUSE 语义韵的对比研究[J]. 现代外语, 2005(3):297-307.
- [19] 梁茂成. 什么是语料库语言学[M]. 上海: 上海外语教育出版社, 2016.
- [20] 熊文新. Web、语料库与双语平行语料库的建设[J]. 图书情报工作, 2013(10):128-135.
- [21] Davies M. iWeb: The 14 Billion Word Web Corpus [EB/OL]. (2018-05-01) [2019-04-05]. <https://www.english-corpora.org/iweb/>.

A Probe into English Vocabulary Teaching Under the Perspective of Data-Driven Learning

XU Liang^{a,b}, ZHANG Yong^a

(a. School of Foreign Languages, Jiangxi Agricultural University, Nanchang 330045;

b. School of Foreign Studies, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: As one of the hot and important topics in foreign language teaching, the corpus-based method of Data-Driven Learning (DDL) is popular in China. DDL is based on authentic language materials, and it represents the principle of "teacher-oriented and student-focused teaching" in which students' inquiry-based learning is formed. By using the example of a pair of English synonym "edition" and "version", the researchers presented the teaching methods and practice of KWIC, N-gram analysis, collocation and colligation, and Corpus on the Web under the perspective of DDL. The study suggests that teachers may construct a small corpus for specific-purpose language teaching by using online English corpus like BNC, COCA, and the iWeb Corpus, which makes it effective in English vocabulary teaching, especially in synonym distinction.

Key words: data-driven learning; corpus; English vocabulary teaching

(责任校对 龙四清)