



当代外国语言文学研究文库

LINGUISTICS

基于语料库的英语语言特征研究

Corpus-based Studies of Linguistic Features in English Texts

张继东 赵晓临 主编

教育部人文社科项目“科技文本中实词基形和屈折变形搭配的异质性研究”(09YJA740019)的成果之一
当代外国语言文学研究文库

基于语料库的英语语言特征研究

——语料库语言学与英语教学研究论集

Corpus Linguistics in English Teaching and Research

主编 张继东 赵晓临
编委 方 燕 李晶洁 刘 晶 刘 萍
Paul Baker (英国兰卡斯特大学)
吴 蕾 杨 捷 张继东 赵晓临

上海交通大学出版社

内 容 提 要

本书是教育部人文社科项目“科技文本中实词基形和屈折变形搭配的异质性研究”(09YJA740019)的成果之一。

全书分为三个章节:第一章主要是从语料库语言学方法论的角度,对通用英语中词汇的“意义单位”所关涉的相关内容进行了详细的描述和分析;第二章集中比较了中国学习者与本族语者在英语使用方面存在的语言特征差异,并对语料库在将来课堂教学中发挥的作用进行了展望;第三章着重描述了科技英语中实词使用的差异性特征。

本书可作为语料库语言学研究人员参考使用。

图书在版编目(CIP)数据

基于语料库的英语语言特征研究/张继东,赵晓临主
编. —上海:上海交通大学出版社,2012

(当代语言学文库)

ISBN 978-7-313-08296-1

I. 基... II. ①张... ②赵... III. 英语—语
言学—研究 IV. H31

中国版本图书馆 CIP 数据核字(2012)第 068550 号

基于语料库的英语语言特征研究

张继东 赵晓临 主编

上海交通大学出版社出版发行

(上海市番禺路 951 号 邮政编码 200030)

电话:64071208 出版人:韩建民

常熟市梅李印刷有限公司 印刷 全国新华书店经销

开本:787mm×960mm 1/16 印张:14.5 字数:270 千字

2012 年 5 月第 1 版 2012 年 5 月第 1 次印刷

印数:1~2030

ISBN 978-7-313-08296-1/H 定价:38.00 元

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系
联系电话:0512-52661481

前 言

《语料库语言学与英语教学研究论集》是教育部人文社科项目“科技文本中实词基形和屈折变形搭配的异质性研究”(09YJA740019)的成果之一,它见证了东华大学语料库语言学研究中心成员历时多年的学术努力。

语料库语言学的研究方法展现了语言学研究中的经验主义实质,体现为实证和量化两个标志性特征。在语料库的支持下,关于语言本质、构成和功能的描写和理论的提升都是在语言实例和数据的基础上做出的客观判断,而不是仅依赖语言学家的语言直觉进行主观臆测。计算机科学的发展使得语言研究工具获得技术的适时升级,语言学家不仅可以凭借数据库平台构建满足研究所需的语料库,也可以使用语料库软件来分析语料、检索和提取语言实例以及数据。自20世纪中后期以来,语料库语言学作为一种新的语言学研究方法发展迅猛,语料库以“非人力所能及”的真实语料检索和数据驱动研究优势在语言学各领域中发挥着“颠覆”传统语言研究方式的作用。目前,语料库语言学正在摆脱工具性地位的束缚,为争取获得与语义学和语用学比肩齐名的学科席位而努力。

在国内语料库的相关研究起步时,张继东老师和赵晓临老师已经开始在此领域进行了深入的探索,并在学校各级领导的支持下,成立了语料库语言学研究中心。两位老师孜孜不倦,持之以恒,带领青年骨干教师和研究生在语料库的基础上对语言特征展开了多维度的研究,并取得了丰硕的学术成果,得到了业界的广泛认可。为了能够系统地梳理相关研究成果,我们将以我院英语教师和研究生为主体的研究人员近年来撰写和发表的论文结集成书,以进一步推动国内的语料库语言学学术研究。本论文集分为三个部分:第一部分主要是从语料库语言学方法论的角度,对通用英语中词汇的“意义单位”所涉及的相关内容进行了详细的描述和分析;第二部分集中比较了中国学习者与英语本族语者在英语使用方面存在的语言特征差异,并对语料库在将来课堂教学中发挥的作用进行了展望;第三部分着重描述了科技英语中实词使用的差异性特征。三个部分的编排体现了我院的语料库语言学研究的学术提升过程。

值得一提的是,在张继东老师的努力下,东华大学语料库研究中心与英国兰卡斯特大学语言学和英语语言系建立了良好的合作关系,该系语料库语言学专家、*Corpora* 杂志的执行主编 Paul Baker 教授欣然为本论文集撰稿,同时该书在组稿的过程中也得到了 Geoffrey Leech 教授、Tony McEnery 教授、Andrew Hardie 博

士的指点和帮助。在此,我们表示衷心感谢。

本书所收学术论文能够集结成册、付梓印刷,离不开多方的支持。在这里,我们由衷地感谢东华大学领导对语料库语言学研究长期的关心和支持,感谢教育部人文社科项目鉴定专家的厚爱鼓励,感谢所有同仁、朋友和学生的参与、支持和奉献!

编 者

目 录

通用英语词汇研究

基于语料库的词汇型式研究·····	赵晓临	(1)
Diachronic Lexical Change in American English (1961—2006) ·····	Paul Baker	(10)
动词 happen、occur 和“发生”的语言差异性探究 ——一项基于英汉语料库的调查与对比分析 ·····	张继东 刘 萍	(28)
英语词汇增幅率研究 ·····	李晶洁 冯志伟	(35)
基于语料库的同义词辨析一般方法 ·····	张继东 刘 萍	(54)
句位:英语同义强势词差异特征研究的新视角·····	董斐娜 张继东	(64)
模糊限制语 somewhat 搭配行为语用功能分析 ·····	董斐娜 张继东	(74)

中国学习者语言特征研究

中国大学生英语书面语中的态度立场表达·····	赵晓临	(81)
语料库研究的课堂教学价值 ——评《从语料库到课堂:语言使用和语言教学》·····	赵晓临 殷 耀	(90)
网络语料库索引行信息在外语教学中的应用·····	赵晓临	(101)
动词突显度和情状体对现在完成时标记影响的诱导性 研究·····	刘 晶 杨 捷 张 琳 张兰芳	(108)
论模糊限制语使用的差异 ——基于 COLSEC 与 COCA 的对比分析 ·····	吴 蕾 张继东	(119)
中国新闻英语的句法修辞特征研究 ——中美事故型新闻报道的文本对比分析 ···	方 燕 杨 捷 张继东	(126)
中国大学生英语写作中的使役结构以及词化现象的调查与 分析·····	张继东 刘 萍	(139)
中国英语学习者强势词句法位置应用的语料库研究 ···	徐 峰 张 振	(149)

科技英语词汇研究

学术英语文本中连续短语单位的提取方法·····	李晶洁 卫乃兴	(157)
-------------------------	---------	-------

基于词形的科技文本词语搭配研究·····	张继东	刘 萍	(168)
科技英语中次技术词的词义研究·····	何 燕	张继东	(182)
基于语料库的科技英语情态动词研究·····	何 燕	张继东	(195)
基于生物科技文本的实词搭配的异质性研究·····	魏洁静	张继东	(203)
基于 JDEST 语料库的学术语篇中报道动词的研究 ·····	文慧静	毕 怀	(212)
科技英语中动词名词化的分类与语篇功能分析·····	张继东	吴 蕾	(219)

基于语料库的词汇型式研究

赵晓临

摘要:本文基于中国学习者口语语料库 COLSEC 和伯明翰大学远程登录英语语料库 BoE 的口语子库,对比分析动词 get 的 GET V-ed 的型式行为。研究表明,中国学习者使用该词汇型式的语义特征不明显,get 具有词汇化特征;而英语本族语者使用该词汇型式具有明显的消极语义特征,get 具有非词汇化特征。此结果对英语口语教学有一定的启示作用。

关键词:型式,型式行为,语义特征,非词汇化

1 引语

近年来基于大数据的语料库语言学研究证实,词汇型式和词相互依赖;某个型式与某些词共现,某个词也依赖一定的型式存在。另一方面,词汇型式与意义密切联系,词的意义因其高频使用的典型型式不同而有所差异,而高频出现在某个型式中的词汇往往共有某一方面的意义(Hunston & Francis 2000: 3)。对中国学习者的词汇型式行为研究也备受关注,濮建中(2003)对中国学习者书面语语料库的调查发现,在相同词汇型式下中国学习者与英语本族语者的词语搭配存在较大差异,影响了学习者语言表达的准确性和地道性。但是,对中国学习者口语中常用词汇型式依然缺少研究,本文就此做初步探讨。本研究以动词 get 为例,调查其词汇型式 GET V-ed 在中国学习者和英语本族语者口语中的型式行为,并探讨调查结果对词汇教学的启示作用。选择动词 get 基于三方面原因。首先,get 是大型语料库中最高频出现的常用动词之一,譬如按词频统计,get 在 BNC 语料库的最高频动词表中列第 7 位、在 FLOB 和 Frown 语料库中列第 10 位和第 8 位、在 LLC 语料库中列第 6 位、在中国学习者语料库 CLEC 和中国学习者口语语料库 COLSEC 中列第 7 和第 5 位^①。其次,由动词 get 构成的词汇型式或多词单位数量大,使用率

^① BNC 语料库词容量 1 亿,其中口语语料 10%,书面语语料 90%;FLOB 和 Frown 语料库均为书面语语料库,词容量均为 100 万;伦敦-朗德英语口语语料库 LLC 词容量 50 万;中国学习者语料库 CLEC 为书面语语料库,词容量 100 万;中国学习者口语语料库 COLSEC 词容量为 100 万。

非常高。再次,动词 get 的行为模式是不同研究领域关注的焦点之一,如语料库研究的搭配行为(collocational behaviour)、认知语言学研究的构式(construction)等都涉及诸如 get 等常用词的高频型式。

2 型式、型式行为的研究与界定

最早使用“型式”(pattern)概念的是英国语言学家 A. S. Hornby。长期从事英语教学研究的 Hornby 认为,语言专家的知识应该为普通学习者和教师服务。基于此,他关注语言使用的规律性特征,上述观点在其代表作《英语的型式与使用指南》(*A Guide to Patterns and Usage in English* (1954))中得到了充分体现。书中不仅提出了“型式”的概念,而且探讨了 25 个动词、4 个名词和 3 个形容词的词汇型式及其语境意义。Francis、Hunston 和 Manning(1996,1998)承继了 Hornby 的“型式”概念,描述英语本族语者高频使用的动词、名词和形容词的词汇型式。但与 Hornby 不同的是,Francis、Hunston 和 Manning 采用 Sinclair 倡导的语料库驱动方法,基于英国伯明翰大学的英语语料库 Bank of English(以下简称 BoE^①)探讨问题。Halliday(2004: 46)认为 Hunston 和 Francis 的研究对象定位为“词汇语法”(lexical grammar),即从词汇的角度处理词汇与语法的混合型式。Hunston 和 Francis(2000)将自己的研究称为型式语法(pattern grammar),即研究高频的词汇搭配,并在此基础上概括搭配的词汇-语法型式。根据 Hunston(2008)的观点,型式可以看作构式的子集(sub-set),但是型式语法与构式语法(construction grammar)的研究方法大相径庭。后者从认知角度研究语言,将理论置于比观察优先的地位;而型式语法强调观察的重要性,它以词汇为出发点,是以词汇为中心的语法,是词汇、语法共居一体的语法,与意义和功能密切相关。

语料库语言学研究从诸多角度探讨词汇型式。Sinclair(1991)研究多词(multi-word)型式,即搭配(collocation)和类连接(colligation)的结合体。Stubbs(2002)探讨词链(chains),研究实义词和功能词的关系,以及词与语法范畴的关系。Biber(1999:990)的词串(lexical bundles)和 Scott(1997:59)的词丛(clusters)研究也完全依赖词频数据,探讨多词连续线性序列(linear sequences 或 n-grams)。

综观文献,多词序列的指称尚无统一的标准术语。本文采用 Hunston 和 Francis 提出的词汇型式及其界定办法,词汇型式指与某个词有规律地相关联并对

^① Francis、Hunston 和 Manning 1996、1998 年主编 Collins COBUILD Grammar Patterns 1 卷、2 卷时,英语语料库 Bank of English(BoE)的词容量为 2.5 亿;本文作者的检索时间为 2008 年 10 月,BoE 的词容量为 4.5 亿。

其意义产生影响的所有词汇的组合。如果一个词的某种组合出现的频率高,或者某种组合取决于一个词,或者与一个词相关的某种组合具有明确的意义,我们就可以认定此组合为该词的词汇型式(Hunston & Francis 2000: 37)。理论上讲,一个词可以出现在多个词汇型式中,但是实际语言使用却表明,一个词的型式屈指可数,一种词汇型式也只选择一定范围的词。例如 Francis、Manning 和 Hunston (1996)对 BoE 语料库中动词 decide 的调查显示,该词与型式 V that、V wh-、V wh to-inf、V to-inf、be V-ed、it be V-ed that 共存;词汇型式 it v-link ADJ that 与 7 组形容词高频共现,分别为 likely、marvellous、important & necessary 等(ibid, 1998: 420)。同时,词汇型式与词的意义之间存在着必然的联系,因为在词汇型式中出现的高频搭配词通常意义比较集中,形成一定的语义特征,由此也决定了该词的意义。词汇型式与语义特征之间的关系,本文称之为词汇的型式行为。例如与词汇型式 it v-link ADJ that 共现的 7 组形容词表达特定的意义,像“likely”组表达(不)可能性,“important & necessary”组表达(不)重要性或(不)必要性。

从以上研究成果不难看出,词与词汇型式是密不可分的。词汇的型式行为为集词的形式与意义于一身,应该成为词汇教学的重要内容之一。

3 分析方法及工具

本文从两个方面探讨动词 get 的词汇型式 GET V-ed 的型式行为。一是描述英语本族语者的词汇型式行为;二是对比分析中国学习者的词汇型式行为与英语本族语者的差异。

学习者语料库采用中国学习者口语语料库 COLSEC,词容量 723,299,语料来源为全国大学英语四六级口语考试的实景录像语料;语料检索采用 Mike Scott 开发的 WordSmith。英语本族语者语料库采用英国伯明翰大学的远程登录(telnet)英语语料库 Bank of English 中的 4 个口语子库(以下简称 BoE),词容量为 62939687,语料来源含有英国和美国的广播节目以及日常口语会话;语料检索采用 BoE 语料库的配套检索软件 LookUp。

首先提取 BoE 和 COLSEC 语料库中 GET V-ed 的索引行,然后按照共现频数、MI 值排序,最后对前 50 个右 1 动词过去分词搭配词进行分析。本文之所以按 MI 值提取搭配词,是因为它提供两个词之间搭配力强弱的信息。MI 为正值,说明两词之间存在相互吸引。MI 值越大,两词共现的概率越高,故搭配力越强;反之,MI 值越小,搭配力越弱。虽然按照 MI 值提取的搭配词可能包含语料库中词频很低的词,但是它是反映短语中两词间联系的紧密程度的有用指标(Kennedy 2008: 23)。

4 英语本族语者 GET V-ed 的型式行为

BoE 中共检索到词汇型式 GET V-ed 12313 索引行, 动词过去分词搭配词检索结果如表 1 所示。

表 1 BoE 中 GET 的右 1 动词过去分词搭配词

(a)按共现频数排序					(b)按 MI 值排序						
1	involved	1235	26	arrested	69	1	birched	12.32	26	roped	9.91
2	married	1162	27	bogged	69	2	reacquainted	12.06	27	teased	9.88
3	used	602	28	tired	68	3	sidetracked	11.36	28	nicked	9.85
4	paid	470	29	left	67	4	slagged	11.15	29	burgled	9.82
5	caught	358	30	addicted	67	5	waylaid	11.10	30	hooked	9.79
6	done	246	31	upset	65	6	acquainted	10.92	31	divorced	9.79
7	lost	233	32	put	64	7	fobbed	10.84	32	busted	9.78
8	started	201	33	taken	63	8	caned	10.82	33	strung	9.75
9	stuck	198	34	shot	62	9	mugged	10.79	34	scripted	9.72
10	killed	167	35	engaged	60	10	clobbered	10.79	35	thumped	9.68
11	fed	161	36	asked	56	11	married	10.72	36	clued	9.58
12	bored	141	37	divorced	55	12	bogged	10.64	37	hassled	9.58
13	hurt	140	38	scripted	53	13	psyched	10.58	38	reimbursed	9.57
14	elected	118	39	promoted	51	14	bored	10.45	39	steamed	9.51
15	called	88	40	kicked	50	15	nabbed	10.45	40	knifed	9.51
16	hit	88	41	treated	49	16	mopped	10.41	41	snowed	9.46
17	mixed	87	42	worried	48	17	lumbered	10.25	42	sucked	9.39
18	excited	82	43	pushed	48	18	shortchanged	10.21	43	chucked	9.28
19	interested	81	44	frustrated	47	19	re-elected	10.17	44	mired	9.28
20	confused	80	45	scared	47	20	booted	10.15	45	caught	9.28
21	told	74	46	annoyed	47	21	hitched	10.12	46	electrocuted	9.23
22	dressed	74	47	knocked	46	22	tarred	10.12	47	bullied	9.22
23	thrown	72	48	broken	46	23	addicted	10.10	48	squashed	9.18
24	changed	69	49	sorted	45	24	riled	10.07	49	annoyed	9.14
25	sent	69	50	carried	45	25	whacked	10.06	50	excited	9.10
总频数					7683						

表 1a 为按共现频数提取的 GET 的 50 个右 1 动词过去分词搭配词,表 1b 为按 MI 值提取的 50 个右 1 动词过去分词搭配词。结果显示上述 100 个动词均为行为动词(Biber 1999: 361)。动词 get 的 GET V-ed 型式的行为特征总结如下。

1) GET 的动词过去分词搭配词具有明显的语义特征。我们可将所有搭配词分为三组。第一组搭配词具有消极语义特征,如表 1a 中的 involved, caught, lost, stuck, killed, fed, bored, hurt, hit, confused, thrown, arrested, bogged, tired, left, addicted, upset, shot, divorced, kicked, worried, frustrated, scared, annoyed, knocked, broken 都表达“陷入困境”或“遭受打击”之意,其频数占总频数的 48.5%;再如表 1b 中的 birched, sidetracked, slagged, waylaid, fobbed, caned, mugged, clobbered, bogged, psyched, bored, nabbed, mopped, lumbered, shortchanged, booted, hitched, tarred, addicted, riled, whacked, teased, nicked, burgled, hooked, divorced, busted, thumped, hassled, steamed, knifed, snowed, chucked, mired, caught, electrocuted, bullied, squashed, annoyed, 表达“遭受打击”之意,这些具有消极语义特征的搭配词在 MI 值排序的前 50 个搭配词中的比例高达 78%。频数数据和 MI 值数据说明,GET 与具有消极语义特征的动词过去分词的搭配力非常强。第二组搭配词显示出与第一组相反的语义特征。表 1a 中有三个搭配词 excited, interested 和 promoted 具有积极语义特征,其词频仅占总频数的 2.8%;表 1b 中只有 excited 一个。这说明 GET 与具有积极语义特征的动词过去分词的搭配力很弱。表 1a 和表 1b 中的第三组搭配词表达中性意义,1a 中有 married, used, paid, done, started, elected, called, mixed, told, dressed, changed, sent, put, taken, asked, scripted, treated, pushed, sorted, carried, 它们的频数占总频数的 48.7%;1b 中此类搭配词有 reacquainted, acquainted, married, re-elected, roped, strung, scripted, clued, reimbursed, steamed, 占 MI 值排序的前 50 个搭配词的 20%。可以说,GET V-ed 具有消极和中性语义特征。

2) 动词 get 在词汇型式 GET V-ed 中呈现非词汇化特征。非词汇化特征指的是词汇失去了自身的语义内容(semantic contents),其意义已融入搭配伙伴(collocational partners)的意义之中,这种语言现象在常用词的使用上尤为突出(卫乃兴 2006)。在词汇型式 GET V-ed 中,其意义和动词过去分词融为一体,如下两例所示:

(1) He did, however, also say he feels strongly about the troops and the innocent victims who will **get caught up** in the conflict, so I guess I'd say...

(BoE: npr/US. Text: SU1—910117)

(2) The leaders went out and sat by themselves — that is, Prime Minister

Netanyahu and Chairman Arafat — and decided they wanted to **get started** sooner than that, or just as soon as they can... (BoE: usspok/US. Text: SU2—15)

上例中 get caught up 和 get started 各自已经成为一个整体,前者表达“被动卷入”之意,后者表达“开始”之意。在这两个词语序列中,语法和词汇融合在一起,表达一个整体意义。

综上所述,英语本族语者使用词汇型式 GET V-ed 时,GET 与具有消极和中性语义特征的动词过去分词粘性强,搭配紧密度高;动词 get 在词汇型式 GET V-ed 中表现出非词汇化特征,它仅传递语法意义,换句话说,GET 和动词过去分词共同组成了一个词汇语法单位,语义上不可分割。

5 中国学习者与英语本族语者的 GET V-ed 的型式行为对比

中国学习者口语语料库 COLSEC 中与 GET 共现频数在 3 以上的动词过去分词搭配词、共现频数及 MI 值见表 2。

表 2 COLSEC 中 GET 的右 1 动词过去分词搭配词

(a)按共现频数排序			(b)按 MI 值排序					
1	married	23	1	married	8.60	9	tired	4.49
2	prepared	8	2	worked	7.18	10	worried	4.34
3	passed	5	3	prepared	6.62	11	improved	3.69
4	bored	3	4	caused	5.85	12	increased	3.20
5	involved	3	5	polluted	5.78			
6	started	3	6	passed	5.37			
7	tired	3	7	involved	5.29			
	总频数	48	8	punished	4.86			

从表 2 可以看出,GET 的 7 个右 1 高频动词过去分词搭配词可分为以下两组:一组是 bored 和 tired,表达“陷入困境”之意,其频数占搭配词总频数的 12.5%;另一组由搭配词 married, prepared, passed, started 构成,表达中性行为。MI 值在 3.0 以上的 12 个动词过去分词搭配词中,有 4 个具有消极语义特征,它们是 polluted, punished, tired 和 worried,占搭配词总数的 33.3%; improved 和 increased 具有积极语义特征,占搭配词总数的 16.7%; married, worked, prepared, 表达中性行为,占 25%。involved 的语境意义比较复杂,稍后讨论, caused 为语法错误,不予讨论。我们下面进一步讨论中国学习者与英语本族语者的差异。

1) 中国学习者使用此词汇型式比率低。频数在 3 以上的动词过去分词搭配

词仅有 7 个。

2) 中国学习者使用 GET 与具有消极语义特征的动词过去分词搭配力不强。在 COLSEC 中, 此类搭配词仅占 12 个 MI 值 3.0 以上的搭配词的 33.3%; 而在 BoE 中, MI 值在 9.1 以上的搭配词的比例高达 78%。

3) 中国学习者使用 GET 与具有积极语义特征的动词过去分词搭配, 表现出与英语本族语者极大差异。如 COLSEC 的检索结果中 GET 与 improved 和 increased 搭配实例如下:

(3) ... miss less chan-changes with than others so that you may have less chances to **get improved** yourself. Second, ... (COLSEC: 030275. txt)

(4) ... that er er chronically speaking er the percentage of female enrollment is **getting increased** and the reason for that I think... (COLSEC: 010019. txt)

在 BoE 语料库中, get improved 和 get increased 不存在, 而 BoE 中的 get excited, get interested 和 get promoted 在中国学习者语料库中没有实例。

中国学习者使用的 GET involved 的语义特征也显示出与英语本族语者的巨大差异。中国学习者使用的词汇型式仅限于 GET involved in n, 但是 2/3 的名词搭配词表达“参与好的活动”之意, 具有积极语义特征。检索示例如下:

(5) ... the social abilities, it is a very good way for them to **get involved** in the social activities. (COLSEC: 020072. txt)

我们又在 BoE 语料库中进一步检索了 GET involved in n 的索引行。观察语境发现, 英语本族语者使用 GET involved in n 型式具有明显的消极语义特征。首先, 名词搭配词大多表达“困难或麻烦之事”, 如 war, skirmish, dispute, minutia 等; 其次句中的谓语动词表达“不情愿”之意, 如 avoid, not want (willing) to, reluctant (reluctance) to 等, 语境意义表明该词汇型式具有“不愿参与费力或困难的活动”之意。

综合上述动词 get 的 GET V-ed 词汇型式行为的对比分析, 中国学习者与英语本族语者的差异可以总结为如下几点。第一, 中国学习者对该词汇型式使用率低, 限于少量动词。第二, 中国学习者使用该词汇型式时虽然表达“从事某行为”, 但是语义特征与英语本族语者差别明显。第三, 英语本族语者使用词汇型式 GET V-ed 时, 动词 get 表现出非词汇化特征, 动词 get 与搭配词的意义融合为一体。

6 结论

本文对常用动词 get 的词汇型式 GET V-ed 进行了对比分析。从研究结果看, 英语本族语者的词汇型式中, 搭配词表达消极和中性语义特征, GET 表现出非

词汇化特征;而中国学习者在这些方面存在缺失和不足。这些问题的主要原因可能存在于英语词汇教学中没有充分重视词与词汇型式,以及词汇型式与意义之间的密切联系。本族语者使用常用动词 get 的词汇型式 GET V-ed 时,其搭配词、搭配词的语义特征、词汇型式的语义趋向等型式行为,应该作为词汇教学的重点,以便引导学生关注语言使用的地道性。笔者建议教师在词汇教学中充分利用英语本族语者的语料,如语料库或开源语料库网站,例如柯林斯在线语料库^①,考虑词汇使用的语境,引导学生养成通过语境观察、总结词汇型式中搭配词的语义特征、概括语义趋向的习惯。

本文虽然只探讨了常用动词 get 的一个词汇型式 GET V-ed 及其型式行为,但它从一个侧面揭示了常用词的词汇型式及其语义特征对语言教学和学习的重要性。当然,如何加强常用词的词汇型式及其型式行为教学,我们还需要进一步深入研究。

参考文献

[1] Francis, G., Hunston, S. and Manning, E. Collins COBUILD Grammar Patterns 1: Verbs [M]. London: HarperCollins. 1996.

[2] Francis, G., Hunston, S. and Manning, E. Collins COBUILD Grammar Patterns 2: Nouns and Adjectives [M]. London: HarperCollins. 1998.

[3] Halliday, M. A. K. & Matthiessen, M. I. M. An Introduction to Functional Grammar (3rd ed.) [M]. London: Hodder Arnold. 2004.

[4] Hornby, A. S. A Guide to Patterns and Usage in English [M]. London: OUP. 1954.

[5] Hunston, S. and Francis, G. Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English [M]. Amsterdam: John Benjamins. 2000.

[6] Hunston, S. Starting with the Small Words: Patterns, Lexis and Semantic Sequences [J]. International Journal of Corpus Linguistics, 2008 (3): 271-295.

① 柯林斯在线语料库(The Collins WordbanksOnline English Corpus Sampler)为词容量 5600 万的开源附码语料库,可以检索英国英语、美国英语书面语,以及英国英语口语语料。网址为 <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>。

[7] Kennedy, G. Phraseology and Language Pedagogy: Semantic Preference Associated with English Verbs in the British National Corpus [C]. In F. Meunier & S. Granger (eds). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins. 2008: 21-41.

[8] Scott, M. *WordSmith Tools* [M]. Oxford: OUP. 1997.

[9] Sinclair, J. *Corpus, Concordance, Collocation* [M]. Oxford: OUP. 1991.

[10] Stubbs, M. Two Quantitative Methods of Studying Phraseology in English [J]. *International Journal of Corpus Linguistics*, 2002 (2): 215-244.

[11] 濮建忠, 英语词汇教学中的类连接、搭配及词块[J], 外语教学与研究, 2003(6): 438-445.

[12] 卫乃兴, Shared Meaning and Delexicalization [R], 第一届国际认知语义学研讨会宣读论文, 2006, 湖南: 湖南师范大学.

A Corpus-based Study of the Pattern of a Word

Xiao-lin ZHAO

Abstract: This paper investigates the behaviour in the pattern GET V-ed, based on College Learner Spoken English Corpus and the Spoken sub-corpora from the telnet Bank of English in University of Birmingham. The investigation indicates that the patterned behaviour displayed by Chinese learners shows no explicit semantic features and GET V-ed has lexical meanings, but by contrast the patterned behaviour displayed by English speakers shows negative features and GET V-ed is non-lexicalized. The findings derive an insight for teaching spoken English.

Keywords: pattern, patterned behaviour, semantic feature, non-lexicalization

Diachronic Lexical Change in American English (1961-2006)

Paul Baker(Lancaster University)

1 Introduction

This paper focuses on the Brown family of corpora, a set of reference corpora, each one million words of size, consisting of 500 text samples, each of about 2000 words, covering 15 genres of published writing. The Brown family retain an important place in the history of corpus linguistics, as its first member was also the first professionally produced reference corpus ever created. Initially referred to as *The Standard Corpus of Present-Day American English*, it consisted of samples from texts that had been published in 1961 and was constructed by W. Nelson Francis and Henry Kucera at Brown University. Eventually, it became known simply as the Brown corpus.

The sampling frame that was created for the Brown corpus was followed in a number of subsequent corpus-building projects, including the LOB (Lancaster-Oslo/Bergen) Corpus, an equivalent corpus of texts published in 1961 in British English, and created in the early 1970s as part of a joint collaboration between Lancaster University, the Norwegian Computing Centre for the Humanities at Bergen and the University of Oslo. Research carried out by Hofland and Johansson (1982) and Leech and Fallon (1992) has compared the Brown and LOB corpora together, focussing on lexical variation and cultural differences.

Since then, new members of the Brown family have been added to both the British and American sides. In the early 1990s, at Freiburg University the FLOB and Frown corpora were created, comprising British and American English published in 1991 and 1992 respectively. Together, these four corpora could be used to examine diachronic change between the 1960s and 1990s, and also compare how such changes differed between British and American English. Some