

基于COCA语料库的study和learn词语搭配研究

闫长红

(陕西理工学院 外国语学院, 陕西 汉中 723001)

【摘要】study和learn为常见的高频词。本文利用美国当代英语语料库COCA—Corpus of Contemporary American English对这两个词的用法,尤其是常用搭配进行了研究,从宏观和微观两方面总结了其典型搭配和总体使用情况,揭示了其使用规律,对语言研究和教学有一定的启发作用和指导意义。

【关键词】COCA语料库;相关度;互信息;词语搭配

【中图分类号】H314 **【文献标识码】**A **【文章编号】**1673-1883(2014)03-0150-03

一、引言

20世纪90年代以来,语言学界掀起了一股运用语料库的热潮[1],语料库已经成为几乎所有语言研究的关键因素(Teubert 2005)。它运用统计学和信息技术,对大量的真实语料进行概率分析,寻找语言使用的规律,具有容量大、语言真实可靠、检索快速准确的特点,为语言学研究提供了一种全新的研究思路,在语言研究和语言教学领域得以广泛应用,尤其是在词汇研究方面,体现出巨大优势。随着最近30年以来二语习得领域对词汇研究的重视,越来越多的学者利用语料库对词汇进行了研究,并取得了较为丰硕的成果。如:濮建忠(2003);陈建生(2004);缪海燕、孙蓝(2005);王海华、陈国华(2007);张继东、刘萍(2008);何华清(2009);胡春雨(2010, 2012)等[2]。

有研究表明[3],中国大学生的英语词汇深度知识还处在较低的水平。因此,提高大学生的词汇能力,尤其是词汇搭配能力成为英语研究和教学的重点。良好的词语搭配知识有助于自然、流利的语言输出,它是外语学习者整体语言能力的重要组成部分(Stubbs 2001),是外语学习的核心内容之一[4]。“词语搭配是在文本中实现一定的非成语意义并以一定的语法形式因循组合使用的一个词语序列,构成该序列的词语相互预期,以大于偶然的几率共现(卫乃兴 2002)”[4]。基于语料库的词语搭配研究最初由Sinclair开始,此后,语料库语言学已形成一套研究词语搭配的方法和手段。

本文利用COCA研究study和learn的使用情况,尤其是它们各自的搭配情况。study和learn为常见的高频词,是表达“学习”这一概念的主要词汇,它们和学生的学习和生活相关度很高,在学生的口语和写作中经常被使用。但是研究表明,学生对这两个词的使用并不准确,比如通常使用learn

knowledge来表达“学习知识”[5],然而在英语语料库中却没有这一表达,而多用acquire/obtain/get/gain learn [6]等。因此对这两个词的研究甚为必要。

二、研究方法

(一)研究问题

study和learn的使用情况,尤其是它们各自的搭配情况。

(二)研究工具

COCA和excel。

美国当代英语语料库COCA—Corpus of Contemporary American English,是由美国Brigham Young University的Mark Davies教授开发的美国最新当代英语语料库,是当今世界上最大的英语平衡语料库,可供免费在线使用(<http://www.americancorpus.org/>)。它规模大(目前词汇高达4.5亿)、语料均衡(涵盖SPOKEN、FICTION、MAGAZINE、NEWSPAPER、ACADMIC五大类型)、速度快、带有词性标注及相关的分析软件,为英语研究者和英语学习者共享美国英语资源提供了一个良好平台。

(三)研究过程

1.在字符串查询区(WORD(S))内输入study. [n*],选择DISPLAY一栏的LIST(列表)选项;同时在SORTING一栏选择RELEVANCE(相关度),然后点击SEARCH(搜索)功能,就可得到按由高到低排列的体现study与相关单词搭配强度的互信息值(MI)列表。

该表精确地体现出所查询的词与其后所搭配词的关系紧密情况。其中Total表示study作动词时与其后1个词距离内所查到的词(如medicine)总数;All表示所查到的词(如medicine)在所有语料库中的总词数;%即表示Total与All的商;互信息MI(mutual information)是对随机的两个词相关性的度量,也就是要查询的词和可能性搭配词在所有语料

收稿日期:2014-06-10

作者简介:闫长红(1974-),男,陕西汉中,讲师,硕士,研究方向:应用语言学、语料库语言学与二语习得。

库中的共现搭配比重(百分比),该比重同时也决定了其互信息的值。如 study medicine 的 MI 值表示 medicine 与 study(作动词)之间的搭配紧密度为 5.60,互信息越高说明二者联合使用的频率越高。

2.在互信息值列表中选出 study 作“学习”时的条目,删掉其它项目(根据牛津高阶英汉双解词典,study 作动词时有“学习,攻读”;“审视,细看”;“研究,调查”等三个义项),以便随后的归纳分析。

3.同 3.1,在字符串查询区(WORD(S))内输入 learn. [n*],选择 DISPLAY 一栏的 LIST 选项;同时在 SORTING 一栏选择 RELEVANCE(相关性),然后点击 SEARCH(搜索)功能,就可得到按由高到低排列的体现 learn 与相关单词搭配强度的互信息值。

4.同 3.2,在互信息值列表中选出 learn 作“学习”时的条目,删掉其它项目(根据牛津高阶英汉双解词典,learn 作动词时有“学习,学会”;“得知,获悉”;“记住,背熟”;“认识到,从…吸取教训”等四个义项),以便随后的归纳分析。

5.在字符串查询区(WORD(S))内分别输入 study. [n*]和 learn. [n*],选择 DISPLAY 一栏的 CHART(图表)选项,然后点击 SEARCH(搜索)功能,就可分别得到 1990-2012 年之间 study 和 learn 在该语料库中出现的总频数及在各个子语料库的使用频数柱状图,有助于对这两词的使用情况有一个宏观了解。

6.把 3.2 和 3.4 及 3.5 的结果分别提取,另存为文本进行相关分析。

三、结果分析

表 1.study 的显著名词搭配词统计数据

排序	CONTEXT	FREQ	ALL	%	MI
44	study physics	26	9978	0.26	6.41
51	study biology	16	7997	0.20	6.03
54	study theology	17	9049	0.19	5.94
57	study engineering	34	18704	0.18	5.89
59	study math	21	12489	0.17	5.78
67	study mathematics	13	8637	0.15	5.62
68	study medicine	40	26920	0.15	5.60
74	study photography	11	8130	0.14	5.46
83	study psychology	16	14551	0.11	5.16
85	study architecture	10	9403	0.11	5.12
88	study economics	11	11259	0.10	4.99
89	study literature	29	29906	0.10	4.98
91	study drama	12	12933	0.09	4.92
94	study philosophy	14	15990	0.09	4.84
98	study accounting	10	11935	0.08	4.77
100	study music	91	118000	0.08	4.65

表 2.learn 的的显著名词搭配词统计数据

排序	CONTEXT	FREQ	ALL	%	MI
1	learn braille	24	3455	0.69	10.25
2	learn English	282	46927	0.60	10.05
3	learn hebrew	11	3125	0.35	9.27
4	learn mathematics	30	8637	0.35	9.25
5	learn math	34	12489	0.27	8.90
7	learn skills	80	46417	0.17	8.24
8	learn languages	13	7552	0.17	8.24
12	learn statistics	14	14373	0.10	7.42
13	learn motor skills	12	13050	0.09	7.34
17	learn techniques	13	19284	0.07	6.89
18	learn French	29	43898	0.07	6.86
19	learn Italian	11	17404	0.06	6.80
20	learn things	154	254487	0.06	6.73
21	learn songs	11	20912	0.05	6.53
22	learn science	35	68492	0.05	6.49
23	learn music	58	118000	0.05	6.43
25	learn language	29	60153	0.05	6.41
26	learn German	11	27049	0.04	6.16
28	learn ways	28	77145	0.04	6.00
30	learn Japanese	10	31758	0.03	5.79
31	learn stuff	14	49766	0.03	5.63
32	learn reading	15	60909	0.02	5.44
33	learn history	31	127103	0.02	5.42
34	learn computer	11	57102	0.02	5.08
35	learn words	15	98516	0.02	4.74

表 1 显示:(1)在互信息值前 100 位中,study 作动词“学习”时的搭配强度主要排在 50 位之后,这表明动词“学习”这一用法在 study 的所有用法中出现频率并不高;搭配强度排在前 50 位的主要是其作名词“学习”时的用法。由此可知,study 表达“学习”这一意义时,主要以名词形式出现。(2) study 作动词“学习”时,其后所跟名词主要有两类:表示“学科”的名词,这占大多数;偶然跟既可表示“学科”,又可表示某种“技艺”的词,如 photography, music。

上表中与 study 搭配的部分词为多义词,在本研究中其具体义项主要通过研读索引行大量例句而判定。

Learn 的显著名词搭配互信息值一共只排到 37 位。表 2 显示:(1) Learn 作动词“学习”时的搭配强度排在前 35 位,这表明“学习”这一用法在 learn 的所有用法中出现频率极高;(2) Learn 作动词“学习”时,其后所跟名词主要有 5 类:表示“语言”的名词,如 braille(布莱叶盲文), English, hebrew (希伯来语),

French等(当然这些词同时也可以表示“学科”),这占大多数;表示“学科”的词,如 mathematics, statistics, history;表示“技能”的词,如 skills, techniques, reading;表示具体对象的词,如 songs, music, computer, words;表示具体对象的词,如 things, stuff, ways。

同时由表1和表2可知,study和learn后均可跟表示“学科”的词,那么要表达学习某个“学科”时,到底选用哪个词?对以上两词的常见搭配进行总

图1 study在各子库及时间段内的柱状分布图

SECTION	ALL	SPOK	FIC	MAG	NEWS	ACA	1990-1994	1995-1999	2000-2004	2005-2009	2010-2012
FREQ	144755	8942	6207	22801	13909	92896	27040	27567	32207	37772	20169
PER MIL	311.75	93.56	68.64	238.61	151.66	1,020.09	260.00	266.48	312.87	370.14	388.52
SEE ALL SUB-SECTIONS AT ONCE											

图2 learn在各子库及时间段内的柱状分布图

SECTION	ALL	SPOK	FIC	MAG	NEWS	ACA	1990-1994	1995-1999	2000-2004	2005-2009	2010-2012
FREQ	60419	10545	8508	15538	10121	15707	12693	13332	13498	13926	6970
PER MIL	130.12	110.33	94.08	162.60	110.35	172.48	122.05	128.88	131.12	136.47	134.26
SEE ALL SUB-SECTIONS AT ONCE											

图1为study一词在五大子语料库中以及五个时间段内的柱状分布图。图中数据共有3行:第一行为各语料库中所查询词语的总词数;第二行为study在各语料库的每百万词频率(PER MIL);第三行柱状图,是依据第二行的每百万词频率(PER MIL)计算显示。图2为learn一词在五大子语料库中以及五个时间段内的柱状分布图。分析以上图表可知:(1)从宏观角度来看,在整个语料库中study比learn更为常用,前者的使用总频率为144755,后者为60419。在各个时间段内也如此;(2)在五大语域,study的使用频率从高到低排序依次为学术文章>杂志>新闻>口语>小说;这和learn的使用情况一致。(3)learn的使用频率在学术文章和杂志之间差别并不是很大(172.48/162.60),在新闻、口语、小说三

种语域之间使用频率差别也很小,尤其是新闻和口语之间近乎相等(110.35/110.33)。而study在学术语域的使用频率远高于其它语域,其它四个语域之间使用频率差别也较大。(4)在学术、杂志、新闻三大语域内,study的使用频率高于learn,尤其是在学术语域;而在口语、小说语域learn的使用频率稍高于study。(5)在各个时间段内,这两个词的使用频率均有逐年递增趋势,但增幅并不大。

结,不难发现:(1)表达学习语言学科时,常用learn;

(2)表达学习“理工科”时,“统计学”,“数学”和“科学”前多用learn;其它学科前多用study;(3)学习“文科”时,多用study;(4)学习“艺术”及艺术类学科时,两者均可以,但learn更常见,因为其搭配互信息值更高(如 study music, MI=4.65, learn music, MI=6.43)。总体来看,表达“学习”这一概念时,learn的使用范围更广,后可跟五类搭配词,而study的使用范围较为狭窄,只能跟两类搭配词。

种语域之间使用频率差别也很小,尤其是新闻和口语之间近乎相等(110.35/110.33)。而study在学术语域的使用频率远高于其它语域,其它四个语域之间使用频率差别也较大。(4)在学术、杂志、新闻三大语域内,study的使用频率高于learn,尤其是在学术语域;而在口语、小说语域learn的使用频率稍高于study。(5)在各个时间段内,这两个词的使用频率均有逐年递增趋势,但增幅并不大。

四、小结
本文利用COCA研究了study和learn的使用情况。语料库为语言教学和研究提供了一种全新的视角和研究思路,它既是一种资源又是一种方法,具有巨大优势。科学恰当地利用语料库进行研究可以使语言研究和教学更加富有成效。

注释及参考文献:

[1]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).
 [2]韩菁菁.近10年来语料库二语词汇习得研究综述——基于国内9种外语核心期刊11年(1999-2009年)的统计和分析[J].江苏科技大学学报(社会科学版),2010(12).
 [3]濮建忠.英语词汇教学中的类联接、搭配及词块[J].外语教学与研究,2003(6).
 [4]邓耀臣,肖德法.中国大学生英语虚化动搭配型式研究[J].外语与外语教学,2005(7).
 [5]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).

(下转156页)

To Explore the Application Type Undergraduate Japanese Listening Classroom Teaching Reform

WANG Fang

(Foreign Language College, Fuzhou College of Foreign Studies and Trade, Fuzhou, Fujian 350202)

Abstract: The traditional single application type undergraduate Japanese listening classroom teaching mode and teaching content behind seriously hindered the Japanese professional talent training. This article explores three aspects of the establishment of "visual three-dimensional teaching model of listening speaking", innovation of listening teaching content, innovation of listening teaching tool and makes suggestions for the application type undergraduate Japanese listening classroom teaching reform.

Key words: listening classroom; teaching mode; audiovisual; teaching content

(责任编辑:周锦鹤)

(上接第152页)

[6]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.

[7]汪兴富,Mark Davies,刘国辉.美国当代英语语料库(COCA)——英语教学与研究的良好平台[J].外语电化教学,2008(9).

A Study of the Collocation of Study and Learn Based on COCA

YAN Chang-hong

(School of Foreign Language, Shanxi University of Technology, Hanzhong, Shanxi 723001)

Abstract: Study and learn are two high-frequency words. This paper studies their usage especially the collocations based on the Corpus of Contemporary American English. The results may have some implications and bear some directive significance for language study and teaching.

Key words: COCA; relevance; MI(mutual information); collocation

(责任编辑:周锦鹤)