

基于语料库的 real 和 true 与名词的搭配用法研究*

邓海东

(广东科贸职业学院 外语系, 广东 广州 510430)

【摘要】利用语料库研究方法,基于美国当代英语语料库(COCA),本文比较分析 real 和 true 的词频、搭配及语义韵。研究表明,real 的使用频率要高于 true,最显著差异体现在报纸语域。True 更倾向于与积极语义韵及中性语义韵名词搭配使用,与 true 搭配的积极语义韵名词集中的 MI 值排序的前列。

【关键词】语料库;词频;搭配;语义韵

【中图分类号】H313 **【文献标识码】**A **【文章编号】**1673-1883(2013)02-0141-04

在 EFL(English as a Foreign Language)学习中,词汇是重点之一。英语的词汇量相当大,其中同义词、近义词占总词汇量的 60% 以上,这些词在词形、词义、结构或用法等的相同或相近给 EFL 学习者造成困惑,因此近义词、同义词辨析是词汇学习和教学的难点。在教学过程中,单纯借助词典要把近义词、同义词辨析往往讲透彻的难度不小;近些年来,语料库的建立和语料库语言学的发展为同义词辨析的研究提供了新的途径。“语料库作为生活语言的抽样可由高级的计算机提供,这将给人们的研究开辟新的途径”。^[1] Sinclair & Renouf 认为学习英语的重点有:语言中的常用词;常用词用法的主要模式;常用词构成的典型搭配。^[2]因此,本文尝试用语料库的研究方法,对常用词 real 和 true 的词频、搭配及其语义韵进行个案研究。

一、研究设计

本文通过对 THE CORPUS OF CONTEMPORARY AMERICAN ENGLISH (COCA) 语料库的检索,统计出 real 和 true 在 COCA 不同语域中词频差异,分析 real 和 true 右侧的第一个名词搭配信息及相应数据,通过分析检索出的搭配特点,归纳 real 和 true 两词在不同的语域、搭配关系及语义韵的异同。

(一) 研究数据

本文数据来源于美国杨伯翰大学 Mark Davies 教授开发的美国当代英语语料库(COCA),单词容量达 4.25 亿,是美国目前最新的当代英语语料库,也是当今世界最大的英语语料库。它包括了 1990 年至 2012 年间口语、小说、新闻、学术文章等广泛的语料资源。因为 COCA 语料库具有容量大、取材广泛和更新快的特点,在常用词的研究中大量使用。

(二) 研究问题

通过分析在 COCA 语料库检索出的 real 和 true 相关数据,试图找出 real 和 true 在不同语域中的使用频率及倾向;与 real 和 true 搭配的名词在语义韵上的区别,与 real 和 true 搭配的典型名词。

二、数据统计与讨论

(一) real 和 true 在语料库的不同语域中的词频分布差异

所谓语域(Register)就是指语言随着使用场合环境不同而区分的语言变体,是指在特定的语言环境中使用的、有一定的语言特征的语言变体。^[3]英国语言学家韩礼德(M. A. K. Halliday)将语域定义为,语言变体可以按照使用的情况划分为语域。了解不同的词在不同的语域的出现频数,观察一词在其母语使用者中更倾向于使用该词而不是其它词,有助于学习者掌握同义词的地道的运用。

本文根据 COCA 提供的频数分布数据,按口语、小说、新闻、学术文章的分类,对 real、true 在不同语域中的分布频数进行分析。以 real 和 true 作为搜索词,在 COCA 中检索的结果如表 1 所示:

表 1 real 和 true 在 COCA 在不同语域中每百万词中出现频率表

	Real	True
口语	407.17	268.78
小说	297.44	197.55
杂志	308.44	189.40
报纸	314.66	135.33
学术文章	204.26	181.03
忽略语域	307.63	195.14

由上表可看出,总体而言,real 出现的频率要高于 true,在忽略语域的情况下,real 的出现频率是 true 的 158%,差距最大的是报纸语域中,real 的出现

收稿日期:2013-03-20

*基金项目:广东科贸职业学院 2013 年度课题“基于 DDL 教学模式对高职高专英语专业学生英语学习得影响的研究”(编号:GDKM2013-13)。

作者简介:邓海东(1976-),男,江西瑞金人,硕士,研究方向:英语教学。

频率为 true 的 233%;而在学术文章语域中差距最小,real 的出现频率仅为 true 的 113%。

real 和 true 两词都在口语语域中出现的频率最大,而两词出现频率最小的语域却不同: real 在学术文章语域中出现频率最小,true 在报纸语域中出现频率才是最小的。

(二)real 和 true 的词语搭配及语义韵差异

通过分析词语搭配是掌握同义词辨析的重要方法之一。词语的搭配作为词汇研究的重点,在语料库建立并发展的基础上,为词汇研究提供丰富的语料资源。从语料库中将节点词的所有搭配检索出来,利用统计方法来计算各搭配词与节点词共现的显著程度。另一方面,搭配词的语义韵也是词汇研究的内容之一。不同语义韵的词汇传递出的是不同的氛围。接下来,本文将就 real 和 true 搭配词的语义韵及典型搭配程度进行分析。

语义韵是语料库语言学近年来一直关注的问题。卫乃兴将语义韵定义为:某些英语词汇的搭配行为显示了一种特殊的趋向,它们习惯性地吸引某一类具有相同或相似语义特点的词与之构成搭配;由于这些具有相同或相似语义特点的搭配词习惯性地、循环往复地与节点词(node word)在文本中共现,节点词也就被染上了相关的语义特点,整个跨距内弥漫着一种特殊的语义氛围,这叫语义韵(semantic prosody)^[4]由此可见,语义韵不仅仅是一种搭配,还涉及到词汇与其搭配词的一致关系,可以超越单个词汇的意义而扩散到整个句子甚至语境,其主要功能是用来表达语言使用者的态度和观点。Stubbs 将语义韵大体分为消极语

义韵(negative prosody)、积极语义韵(positive prosody)和中性语义韵(neutral prosody)三类。^[5]本文在将跨距(span)设置为节点词右侧第一个名词(-0/+1)的条件下,检索 real 和 true 在 COCA 中的口语、小说、新闻及学术文章四大语域的词语搭配来分别对 real 和 true 进行语义韵及词语搭配描述。

表 2 在 COCA 语料库中 real、true 右侧搭配的名词(Top 50)的语义韵频率及频数分布表

		积极	中性	消极
Real	频数	11	27	12
	频率	22%	54%	24%
True	频数	14	30	6
	频率	28%	60%	12%

据表 2 可知,在共现频率排序的前 50 个与 true 及 real 右边搭配的第一个名词中,最显著的区别是与 true 搭配的消极语义韵名词仅为 real 的一半,与 true 搭配的积极语义韵、中性语义韵名词比 real 略多。由此可见,与 real 相比较而言,true 更多地与积极语义韵和中性语义韵名词搭配使用。

语义韵只是从类别层次上的分析,为了进一步分析 real 和 true 与不同的搭配词间的相关性,将常用的测量值之一的 MI 值纳入分析中是很必要的。MI 值(Mutual Information Score, 互信息值)是指互相共现的两个词中,一个词对另一个词的影响程度或者说一个词在语料库中出现的频数所能提供的关于另一个词出现的概率信息。MI 值越大,说明节点词对其词汇环境影响越大,对共现词的吸引力越强。简而言之,MI 值表示的是词语间的搭配强度。

表 3 在 COCA 语料库中 real、true 右侧搭配的名词(Top 50)频率及 MI 值分布表

Real			True		
搭配词	共现频率(%)	MI 值	搭配词	共现频率(%)	MI 值
ESTATE	61.74	10.97	BELIEVERS	13.75	9.46
EYE-OPENER	21.8	9.47	BELIEVER	11.85	9.25
HOUSEWIVES	10.8	8.46	BETRAYALS	11.11	9.15
LOOKER	6.17	7.65	GRIT	8.32	8.74
CULPRITS	6	7.61	CONFESSIONS	3.64	7.55
STUNNER	5.28	7.42	SELVES	3.45	7.47
CULPRIT	5.13	7.38	WOMANHOOD	3.31	7.41
SHOCKER	5.05	7.36	VOCATION	2.82	7.18
DOWNER	3.58	6.86	POSITIVES	2.32	6.89
WAGES	3.45	6.81	NATURES	1.72	6.47
GNP	3.3	6.75	FRESHMAN	1.48	6.24
GDP	3.04	6.63	GREATNESS	1.3	6.06
DANGER	2.7	6.46	FRESHMEN	1.25	6
CHARMER	2.63	6.42	TEMPER	1.24	5.99

BREAKTHROUGH	1.99	6.01	SELF	1.24	5.99
CRUX	1.93	5.97	PANTRY	1.08	5.79
POSSIBILITY	1.82	5.89	INTENTIONS	1	5.68
INCOMES	1.58	5.68	COLORS	0.93	5.58
BOON	1.57	5.68	MEANING	0.84	5.44
KICKER	1.56	5.66	CONFESSION	0.81	5.38
KNACK	1.53	5.64	FRIENDSHIP	0.78	5.32
ACADEMIA	1.52	5.63	NEGATIVES	0.77	5.3
THREAT	1.5	5.61	STORY	0.76	5.28
DANGERS	1.5	5.6	ROMANCE	0.75	5.27
WAKE-UP	1.48	5.59	HAPPINESS	0.75	5.26
PROBLEM	1.44	5.55	FEELINGS	0.74	5.25
LOSERS	1.43	5.54	CULPRIT	0.73	5.23
HOOT	1.43	5.54	NATURE	0.7	5.17
DISSERVICE	1.42	5.53	HEROES	0.7	5.16
INROADS	1.41	5.52	COMPASS	0.69	5.15
VILLAIN	1.4	5.51	HEIRS	0.68	5.13
PAYOFF	1.38	5.49	DIARY	0.67	5.11
HEROES	1.37	5.47	IDENTITY	0.67	5.1
BATTLEGROUND	1.34	5.45	EMERGENCIES	0.64	5.05
KILLER	1.34	5.44	PATRIOT	0.56	4.83
CHALLENGE	1.33	5.44	GENIUS	0.55	4.83
REFORMER	1.31	5.42	LOVE	0.55	4.82
NAME	1.3	5.41	PATRIOTISM	0.55	4.81
MURDERER	1.26	5.36	HEIR	0.52	4.74
DILEMMA	1.24	5.33	GENTLEMAN	0.51	4.71
TRAGEDY	1.23	5.32	TALES	0.51	4.7
WORLD	1.2	5.29	REFORMERS	0.5	4.69
MOTIVE	1.2	5.29	INCIDENCE	0.5	4.69
KILLERS	1.18	5.27	MOTIVES	0.5	4.68
GEMS	1.17	5.25	EQUALITY	0.49	4.66
LIFE	1.16	5.24	ESSENCE	0.49	4.65
THRILL	1.13	5.2	PREVALENCE	0.48	4.63
JERK	1.11	5.18	IDENTITIES	0.48	4.62
REASON	1.1	5.16	PIONEERS	0.45	4.52
GEM	1.05	5.09	SUPERSTAR	0.44	4.5

由表3可知,据MI值等于或大于3的词作为显著搭配词,real和true右侧搭配的第一个名词MI值最大的前50个名词都属于显著搭配词。与real搭配强度最大的是中性语义韵名词estate,在COCA语料库中共现频率高达61.74%;与true搭配强度最大的则是积极语义韵名词believer(s),共现频率为25.6%,其中排序第一的是believers,共现频率为13.75%,排序其次的是believer共现频率为11.85%。

在与true搭配的名词中,MI值排序第三是消极语义韵名词betrayals,共现频率11.11%,MI值为9.15。与real搭配强度最大的消极语义韵名词是culprits,MI值7.61,排序为第五。纵观MI值排序前10的名词中,与real搭配的有4个为中性语义韵,积极语义韵及消极语义韵的分别有3个;与true搭配的中性语义韵、积极语义韵各有4个,消极语义韵为2个。在MI值排序前10的范围内,与real搭配的名

词排序前5中都没有消极语义韵的词,而与 true 搭配的名词排序前5没有中性语义韵的词。

在MI值排序前50的名词中,与 real 和 true 搭配的词各有特点。与 real 搭配的11个积极语义韵名词中,有8个出现在MI值排序前25的范围内;12个消极语义韵名词中,有7个集中在MI值排序范围为第21-31的区间内。与 true 搭配的14个积极语义韵名词中,有7个出现在MI值排序前25的范围内,另有4个集中在MI值排序为第35-40的区间内;6个消极语义韵名词中,有2个在MI值排序第3-5的区间,另有3个在MI值排序第20-27的区间内。

三、结语

通过以上分析,我们可以看出:real 和 true 在适用语域、搭配名词的语义韵类别及其分布上都存在差异。

首先,在忽略语域的情况下,real 出现的频率要比 true 高出58%,尽管 real 和 true 两词出现频率最高的都是口语语域中,出现频率最低的语域却不同:

real 是在学术文章语域,true 是在报纸语域。这显示出同义词出现的频率与语言系统的倾向程度具有相关性,也揭示了同义词在实际运用中具有语域优先情况。

其次,在共现频率排序的前50个与 true 及 real 右侧搭配的第一个名词中,中性语义韵名词都超过50%,但与 true 搭配的消极语义韵名词仅为 real 的一半。与 real 相比较而言,true 倾向于与积极语义韵及中性语义韵名词搭配使用。

第三,在与 real 和 true 右侧搭配的第一个名词 MI 值最大的前50个名词中,MI 值排序第一分别是中性语义韵和积极语义韵的名词,与 real 搭配的消极语义韵名词集中分布在MI值排序第21-31区间的中间位置,积极语义韵名词的分布较分散,与 true 搭配的积极语义韵名词集中分布在MI值排序靠前第1-4区间和偏后第35-40区间,消极语义韵名词的分布较分散。了解以上差异,将有助于更好地掌握 real 和 true 两词,更好地提高英语综合能力。

注释与参考文献:

- [1] SINCLAIR J. Corpus Concordance Collocation [M]. Shanghai: Shanghai Foreign Language Education Press, 1999. 23.
- [2] Sinclair, J. & Renouf, A. A Lexical Syllabus for Language Learning [A]. In Carter & McCarthy (eds.). Vocabulary and Language Teaching [C]. Longman: Longman, 1988. 148-149.
- [3] 侯瑞德. 英语语体 [M]. 上海: 上海外语教育出版社, 1999: 9.
- [4] 卫乃兴. 语义韵研究的一般方法 [J]. 外语教学与研究, 2002, 4.
- [5] Stubbs, M. Text and corpus analysis [M]. Oxford: Blackwell Publishers Ltd, 1996. 176.

A Corpus-based Analysis on the Noun Collocation of “real” and “true”

DENG Hai-dong

(English Department, Guangdong Vocational College of Science and Trade, Guangzhou, Guangdong 510430)

Abstract: With corpus linguistics research method and based on the statistic from The Corpus Of Contemporary American English (COCA), it makes a comparative study on word frequency, collocation and semantic prosody between “real” and “true” in this paper. The research indicates that “real” is more often used in all register than “true”, especially in newspaper. “True” prefers to modify nouns with positive and neutral prosody, and the positive prosody nouns modified by “true” rank in first forefront according to Mutual Information Score.

Key words: Corpus; Word Frequency; Collocation; Semantic Prosody

(责任编辑:周锦鹤)