

基于语料库的高频形容词 different 词汇知识研究

闫长红

(陕西理工学院 外国语学院, 陕西 汉中 723001)

【摘要】利用语料库研究中国英语学习者对高频形容词 different 的使用情况,并通过和英语母语者对比,探究两者对该词的搭配能力运用的异同,从而了解中国英语学习者对该词词汇知识掌握的深度,分析了主要原因,并提出了教学启示。

【关键词】语料库;中国英语学习者;词汇深度;词语搭配

【中图分类号】H314.2 **【文献标志码】**A **【文章编号】**1673-1883(2015)03-0058-03

词汇是语言的基本成分,对词汇知识的学习是语言学习的重要内容和基本任务之一。

词汇知识可从深度和广度两方面考察。词汇深度是词汇的“质”的研究,其基本问题是:what is meant to know a word?其基本假设是:Knowing a word is more than just knowing its form and meaning (Richards 1976、Nation1990)^[1]。Nation将词汇知识看成是由不同的成分组成的多维体系,将词汇知识划分为形式、位置、功能和意义四大层面以及发音、书写、句法特征、搭配、使用频率、得体性、概念意义、词间联想八个类型;且每个词汇知识类型又分“接受”和“产出”两种能力^[2]。形容词表示人或事物的性质、状态,属性等,对形容词的运用可以反映语言使用者的思维能力。本文利用语料库研究中国英语学习者对高频形容词 different 的使用情况,并通过和英语母语者对比,探究两者对该词的搭配能力运用的异同,从而了解中国英语学习者对该词词汇知识掌握的深度。

一、研究方法

(一)在 WECCL 中考察 different 作定语时其后的搭配情况

利用 AntConc 的词频排序功能在 WECCL(中国学生英语笔语语料库)中选出 different(排在形容词第三位,总频数为 1710)。点击 Collocate(搭配词),然后在 Search Term 中键入 different;在 Window Span(搭配词跨距)中设置搭配词距离检索词的位置,本研究设置为从 0L 到 1R(即研究 different 右边的第一个搭配词,左边为 0 词,下同)。然后在 Min. Collocate Frequency(最小搭配频数)中设置为默认值 10;最后点击 Start 选项卡,显示出 different 后的所有搭配。其中和名词搭配类型共有 20 项,总频数为 922,选出频数在 20 次以上的语项进行研究。

(二)在 COCA 中研究 different 作定语时其后的搭配

情况

查询 different 作形容词时在 COCA(美国当代英语语料库)中出现的总频数。在字符串查询区(WORD(S))内输入 different,选择 DISPLAY 一栏的 LIST(列表)选项,然后点击 SEARCH(搜索)功能,就可得到总频数 186649。在字符串查询区(WORD(S))内输入 different,点击 Collocate(搭配)并在方框内输入 [n*]。在其后的左右两个方框分别选 0 和 1(即研究 different 右边的第一个搭配词,左边为 0 词,下同),选择 DISPLAY 一栏的 LIST(列表)选项;同时在 SORTING 一栏选择 FREQUENCY(频数),最小值为 10,然后点击 SEARCH(搜索)功能,就可得到频数由高到低排列的前 100 位(最低频数为 189)与 different 搭配的名词列表(总频数为 55819),把其中和步骤 1 中对应的各项选出。

以上两步的结果如表 1 所示。

(三)在 COCA 中分别研究表 1 中 different 后各词与定语形容词搭配情况

在字符串查询区(WORD(S))内输入 people,点击 Collocate(搭配)并在方框内输入 [j*]。在其后的左右两个方框分别选 1 和 0,选择 DISPLAY 一栏的 LIST(列表)选项;同时在 SORTING 一栏选择 FREQUENCY(频数),然后点击 SEARCH(搜索)功能,就可得到频数由高到低排列的前 100 位修饰 people 的形容词列表(总频数为 98881),把其中和 different 意思相近的频数靠前的搭配选出。按照同样步骤,分别研究其它各词搭配情况。

(四)在 WECCL 中分别研究表 1 中的各词与定语形容词搭配情况

点击 Collocates(搭配词),然后在 Search Term 中键入 people;在 Window Span(搭配词跨距)中设置搭配词距离检索词的位置,本研究设置为从 1L 到 0R。然后在 Min. Collocate Frequency(最小搭配频

收稿日期:2015-05-24

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

数)中设置为默认值20;最后点击Start选项卡,显示出people前的所有搭配,把其中和different意义相近的形容词搭配选出。按照同样步骤,分别研究其它几个词的搭配情况。

以上两步的结果如表2所示。

二、结果与讨论

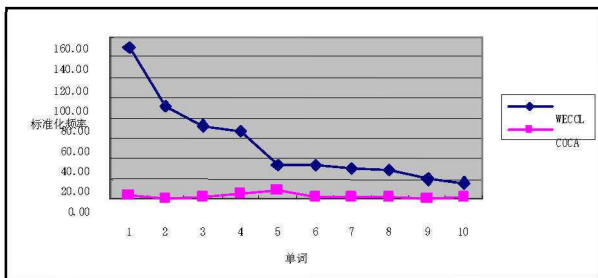
(一)在WECCL和COCA中different作定语时其后的搭配情况

表1 在WECCL和COCA中different作定语时其后的名词搭配频数结果

	1	2	3	4	5	6	7	8	9	10
WECCL Col	people	opinion(s)	countries	kinds	ways	culture(s)	views	places	ideas	areas
Fre	193	117	94	86	43	42	39	36	26	20
Nfre	149.70	90.75	72.91	66.70	33.35	32.56	30.25	27.92	20.17	15.51
COCA Col	people	opinions	countries	kinds	ways	culture(s)	views	places	ideas	areas
Fre	1552	238	777	2366	3777	962	514	860	345	586
Nfre	3.45	0.51	1.67	5.10	8.14	2.07	1.11	1.85	0.74	1.26

(上表中各英文缩写代表内容如下相对排序: Col 搭配; Fre 频数; Fre% 频数百分比; Nfre 标准化频率。以下各图表同)。标准化频率用以比较不同大小的语料库中某个(些)词的出现频数。频数标准化用某个(些)检索项的实际观察频数除以总体频数(通常为文本或语料库的总词数)。再次基础上乘以1千(1万、1百万)得到平均每千(万、百万)词的出现频率。本研究采用每百万词计算,即:标准化频率(每万词) = (观察频数/总体频数) × 1000000^[3]。

图1 在WECCL和COCA中以上单词标准化频率对比



由上表1可知:

(1) 在 1289303 个词的 WECCL 语料库中, different 共出现频数为 1710 次, 占整个语料库比例的 0.13%, 标准化频率为 1326.30。其中作前置定语和名词的搭配类型(频数在 20 以上)共有 10 项, 占 different 总用法的 0.58%(10/1710); 该 10 项总频数为 696, 占 different 使用总频数的 40.7%(696/1710)。在这 10 项搭配中, 频数主要集中于前两项 people 和 opinion(s)。(2) 在容量为 464,020,256 个词的 COCA 语料库中, different 作形容词时总频数为 186649, 占整个语料库比例的 0.04%, 远低于 WECCL 中的 0.13%;

标准化频率为 402.24, 同样远低于 WECCL 中的 1326.30。其中作前置定语和名词的搭配类型(频数在 97 以上)共有 100 项, 占 different 总用法的 0.05%(100/186649), 远低于 WECCL 中的 0.58%; 该 100 项总频数为 55819 次, 占 different 使用总频数的 29.90%(55819/186649), 远低于 WECCL 中的 40.76%。(3) 和 WECCL 相比, COCA 中各词的频数大小排序不同。中国学习者使用频数最多的 people 和 opinion(s) 分别排在母语使用者的第三位和末位; 两者使用频率最高的前两个词的频数远高于其它项, 表明两者对 different 和名词的搭配都集中于前两个词。同时根据图 1 标准化频率对比显示, 中国学习者对以上各搭配的使用频率远远高于母语者, 属于超用。

综上所述: 母语者对 different 的使用频率远低于中国学习者, 也就是说中国学习者超用了 different。母语者对 different 作形容词时的使用范围远远广于中国学习者, 仅频数在 189 以上的搭配就有 100 项之多, 而中国学习者却只有 10 项, 亦即中国学习者对该词的使用范围较为狭窄, 词汇深度远远不够。

(二)在COCA中different后各词与定语形容词搭配情况

表2 在COCA和WECCL中different后各词前定语形容词搭配频数排序结果

	Word	COCA		WECCL					
		Fre	NFre	Word	Fre	NFre			
people	different	1549	3.34	various	246	0.55	different	130	100.83
countries	different	777	1.67	various	254	0.56	different	90	69.81
kinds	different	2336	5.03	various	691	1.54	different	84	65.15
ways	different	3775	8.14	various	888	1.97	different	43	33.35
views	different	514	1.11	differing	153	0.34	different	38	29.47
places	different	860	1.85	various	252	0.56	different	35	27.15
culture(s)	different	794	1.71	diverse	165	0.37	different	42	32.58
				various	132	0.28			
areas	different	586	1.26	various	186	0.41	different	20	15.51
ideas	different	345	0.74				different	26	20.17

对上表以分析可知: 在修饰以上各名词表达“不同的”这一概念时, 母语者和中国英语学习者所使用的频数最高词都是 different, 亦即中国英语学习者在该用法上和母语者较为一致, 用法恰当。但是在标准化频率方面中国英语学习者远高于母语者, 属于超用。这表明母语者能用来和上述名词搭配的形容词范围很广, 可以用更为详细具体的词而不是笼统的 different 来对名词修饰, 比如对 people 就用表

达不同国籍、年龄、肤色、身份等的词来修饰；而中国英语学习者受思维所限，能用的就极少，因而在有限的表达范围内不断重复，使用频率超高。

同时可以看出对于某些名词，不论是母语者还是中国英语学习者，所能用来修饰它们的意义相近的形容词都很有有限，这也许表明了不同民族思维的一致性。

三、原因分析与教学启示

中国英语学习者对词汇知识掌握的深度不够与其词汇学习策略以及语言输入有关。大多数学生学习单词主要是靠词汇表或查词典，这种方法脱离语境，对词汇知识的学习支离破碎，难以有一个较为全面的深刻的把握。可理解性语言输入是语言习得的必要条件，输入的数量和质量对学习者的语言发展意义重大。如前所述，多数学生主要是靠词汇表或查词典而不是在具体的语境中学习词汇，这样的输入数量不足、形式单一，很难提高自己的语言水平。

上述问题可以通过合理利用语料库来解决。

语料库拥有大量真实的语言材料，可以保证语言输入的数量和质量。在课堂中可以引入数据驱动学习法(Data-driven learning,简称DDL)，这是语料库应用于语言教学中的一种教学模式(Johns, 1991)。其主要思想是指引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、自我发现语

法规则、意义表达及语用特征^[1]。使用语料库的检索工具，探究本族语者的词汇搭配情况，从而建立词汇搭配及使用频率的概念。语境共现可以在同一个界面上，以检索词居中的方式呈现该词在一个或若干个文本中出现的所有次数及具体语境，便于学习者集中比较、理解词汇的多项词义，领悟其使用特征及语境。学生通过阅读语境共现探究目标词语与语境中其它词语的联系，总结其搭配规律，完成词汇的意义建构，拓展词汇知识，提高词汇学习的效率。

四、结论

本文利用语料库研究了中国英语学习者对高频英语形容词different的使用情况。总体来看，中国英语学习者对该词的使用范围较为狭窄，作定语时所修饰的频数较高名词只有10个，而英语母语者多达100个以上；大量重复使用该词，且频率超高；在表达“不同的”这一概念时，英语母语者所使用的形容词多于中国英语学习者，并且在很多情况下会使用更为详细具体的词，而中国英语学习者则仅仅局限于“different”，表达较为含糊。这反映了中国英语学习者词汇知识深度不够，并且部分搭配不够恰当。因此教师在教学中可以适当引入语料库，给学生提供大量真实的语料，引导学生自主探索、发现、总结语言规律，丰富词汇知识，提高语言水平。

注释及参考文献：

- [1]肖善香,刘绍龙.论二语词汇深度习得及其研究的若干问题[J].暨南学报(哲学社会科学),2003(1).
- [2]Nation, L. S. P. Teaching and Learning Vocabulary [M]. New York: Newbury House Publishers, 1990:159-164.
- [3]梁成成,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.
- [4]程春梅.对“语料驱动学习”模式的研究[J].国外外语教学(FLTA),2007(3):58-64.
- [5]何安平.语料库辅助英语教学入门[M].北京:外语教学与研究出版社,2010.
- [6]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 Corpus-based Study on Vocabulary of different

YAN Chang-hong

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

Abstract: This paper carries a corpus-based study to explore the Chinese English learners' use of high-frequency adjectives different by contrast with that of native English speakers to probe into their vocabulary knowledge of this word.

Key words: corpus; Chinese English learners; vocabulary depth; collocation

(责任编辑:周锦鹤)