

## 基于语料库的“V (a)round”构式搭配 关联强度及动词“V”语义聚类研究

杨莎莎

(广西大学外国语学院,南宁 530004)

**摘 要:** 基于 COCA 语料库,对出现于“V (a)round”构式的动词进行检索,并使用 Fisher 精确检验计算搭配强度。研究发现:进入构式 V 槽的动词主要有两大类,相对静止动词和位置发生变换的动词;根据动词的不同,构式的凸显有所不同,主要有:对动作结果的突出,对方式、路线和范围的突出,对相邻、相关内容的突出,对动作持续性和连续性的突出;根据构式凸显的不同,构式义相应地可以有:表示动作的结果,表示动作的方式、路线和范围,表示动作相邻、相关内容,表示动作发生时伴随的状态。基于语料库的“V (a)round”构式搭配强度及其语义的分析,不仅有助于进一步了解该构式,也是对构式搭配分析法的进一步应用,对我国外语教学也有一定的启示。

**关键词:** 构式搭配强度;“V (a)round”构式;构式义;构式搭配分析法

**中图分类号:** H059

**文献标志码:** A

**文章编号:** 1673-3851 (2017) 04-0336-06

Goldberg<sup>[1]1-21,65-96</sup> 的构式语法认为,构式是语言的基本单位,构式之间通过各种承继关系构成网络,实现意义的拓展。在构式的相关研究中,构式搭配强度一直为学界所关注。本文主要关注获得构式允准进入“V (a)round”构式中的动词与构式的搭配强度,分析此类动词的语义聚类。

近年来关于各类构式的研究最主要有两大类:一类关注构式内部各项与构式之间的互动关系,例如王仁强等<sup>[2]</sup>、张建理等<sup>[3]</sup>、王寅<sup>[4]</sup>、魏晓龙等<sup>[5]</sup>、方强等<sup>[6]</sup>的研究;另一类则主要关注构式的认知机制及其语义和语用特征,例如许红菊<sup>[7]</sup>、杨江锋<sup>[8]</sup>、袁红梅等<sup>[9]</sup>、席建国等<sup>[10]</sup>的研究。这些研究为各类构式研究提供了良好的认知基础,使得构式理论不断发展和完善,并拓展到更多的语言领域,如汉语(杨永龙<sup>[11]</sup>、张建理等<sup>[12]</sup>、张克定<sup>[13]</sup>、吕建军<sup>[14]</sup>、甄珍<sup>[15]</sup>等)。但此类研究多是通过内省的方法探寻构式意义和用法。随着语料库语言学的兴起,一些学者开始尝试用语料库的方法研究构式,并取得丰富成果(王仁强等<sup>[2]</sup>、罗思明

等<sup>[16-17]</sup>、陈佳<sup>[18]</sup>、Stefanowitsch 等<sup>[19]</sup>)。但基于语料库的研究方法对搭配强度的分析仍十分有限,且主要是对于词汇搭配强度的分析,如 Church 等<sup>[20]</sup> 和 Dunning<sup>[21]</sup> 分别用 *t*-score 模型和 Chi-square 模型计算词汇搭配强度,较少涉及词汇与构式之间的搭配强度,且所运用的数据统计方法也可进一步改进。Stefanowitsch 和 Gries<sup>[19]</sup> 的构式搭配分析法为探寻词汇与构式之间的搭配强度提供了更切实可行的方法,但直到目前为止仍只是被少数研究者使用(杨晶等<sup>[22]</sup>、胡健等<sup>[23]</sup>、田臻等<sup>[24]</sup>、陈佳<sup>[18]</sup>),其广度和深度都远远不够。构式搭配分析法不但可以运用到更多的构式分析中,其方法也在不断地改进。

因此,本文以“V (a)round”这一常用构式为研究对象,使用 COCA 语料库提供的语料,运用构式搭配分析法探讨动词与构式的搭配强度,分析“V (a)round”构式中动词的语义聚类,进一步探讨该构式的构式义。

## 一、构式搭配分析法

构式搭配分析法由 Stefanowitsch 和 Gries<sup>[19]</sup> 于 2003 年首次提出,其目的在于分析比较构式槽位中词素与该构式的关联强度,从而探寻构式意义。该分析法主要利用四格交叉表,通过 Fisher 精确检验计算出每一个共现词素与该构式的显著性  $p$  值。使用 Fisher 精确检验计算搭配强度与  $z$ -score、 $t$ -score 模型或 Chi-square 模型相比的优势在于: $z$ -score、 $t$ -score 模型或 Chi-square 模型均要求所计算的数据呈卡方分布,且要求计算样本足够大,如果四格表中任意一格数值较低或为零,则会影响  $p$  值的准确性;而 Fisher 精确检验则是基于超几何分布(hyper geometric distribution)的统计学原理,对于数据分布的正态性与样本量的大小没有限制,因此,从对文本要求的角度考虑,按照该方法计算得出的显著性  $p$  值的准确度要高于按照  $z$ -score、 $t$ -score、Chi-square 等方法得出的值。<sup>[18]</sup>

针对不同的研究目的和对象,Stefanowitsch 和 Gries<sup>[19]</sup> 的构式搭配分析方法呈现不同的变化:a)共现词素分析法(collexeme analysis),用于计算某个槽位上的词素与该构式的搭配强度;b)(多项)显著共现词素分析法(multiple distinctive collexeme analysis),用于分析出现于某一个槽位的共现词素的不同语义聚类,比较两个或多个看似同义的“可替换对”的构式义;c)互为变化的共现词素分析法(covarying collexeme analysis),用于计算互为变化的共现词素组合与该构式的关联强度。

而关于共现词素分析法,Stefanowitsch 和 Gries<sup>[19]</sup> 指出,欲计算某特定词项  $L$  与某一特定构式  $C$  之间的构式搭配强度,需首先知道四个频数:词项  $L$  在构式  $C$  中出现的频数  $a$ ;词项  $L$  在所有其他构式中出现的频数  $b$ ;构式  $C$  与除词项  $L$  外其他词项共现的频数  $c$ ;构式  $C$  外所有其他构式出现的频数  $d$ 。此四项频数可被做成 4 格交叉表(如表 1),代入 Fisher 精确检验进行计算。Stefanowitsch 和 Gries 以出现在[N waiting to happen]构式 N 槽位中的词素 accident 为例进行了具体计算和分析,计算出“accident”与[N waiting to happen]的搭配强度。在此基础上,为进一步分析该构式 N 槽位的语义聚类,对出现在 N 槽位中的每一个名词进行搭配强度的计算,并按搭配强度的高低进行排列,分析其语义聚类。发现出现在[N waiting to happen]构式 N 槽位中的往往是“accident, disaster, earthquake, invasion”等具消极含义的词素。

表 1 共现词素“accident”与[N waiting to happen]构式的交叉表

构式 词项	accident	非 accident	每行总数
[N waiting to happen]	$a$	$c$	$a+c=M$
非[N waiting to happen]	$b$	$d$	$b+d=N$
每列总数	$a+b=X$	$c+d=Y$	$W=X+Y=M+N$

Stefanowitsch 和 Gries<sup>[19]</sup> 的共现词素分析法为分析词素与构式的关联强度以及进一步探寻构式义的研究提供了可行的研究方法,该计算方法推广到其它构式同样有效。因此,通过 Stefanowitsch 和 Gries 的共现词素分析法,“V (a)round”构式 V 槽位的词素与该构式的搭配强度同样可以得到计算,从而便于该构式构式义的深度挖掘。

## 二、研究方法

运用 Stefanowitsch 和 Gries 的构式搭配分析法,考察出现于“V (a)round”构式 V 槽位的词素与该构式的搭配强度,本研究尝试回答以下 3 个问题:“V (a)round”构式 V 槽位吸引的词素有哪些?“V (a)round”构式 V 槽位的语义聚类是什么?“V (a)round”构式搭配强度分析对外语教学有何意义?

本文主要研究资料 and 工具如下:a) COCA 语料库(Corpus of Contemporary American English)。该语料库由美国 Brigham Young University 的 Mark Davies 教授开发,最大的特点是信息更新快,时效性强。该语料库每 6~9 个月更新一次数据,每年收集的数据约 200 万词,截止 2015 年共包含 4.50 亿单词,涵盖美国 1990 年至今的口语、小说、杂志、报纸和学术期刊 5 大类的语料。本文借助该语料库分别对“V around”以及“V round”构式进行检索。b) SPSS22.0 (Statistical Product and Service Solutions)。SPSS22.0 为 IBM 公司推出,具有完整的数据输入、编辑、统计分析、报表、图形制作等功能。自带 11 种类型 136 个函数。本研究主要利用该软件的交叉表格进行 Fisher 精确检验。

主要研究步骤包括:a)相关频数检索。通过 COCA 语料库获得进行 Fisher 精确检验所需的四项频率  $a$ 、 $b$ 、 $c$ 、 $d$ 。由于四项频数中只有频数  $a$  可以利用检索式直接获得, $b$ 、 $c$ 、 $d$  则需要通过借助该词项在所有构式中出现的总频数  $X$ ,该构式出现的总

频数  $M$ , 所有构式出现的总频数  $W$  经计算后获得 (见表 1)。因此, 本研究需要借助 COCA 语料库提取的数据为: 词项在构式中出现的频数  $a$ ; 该词项在所有构式中出现的总频数  $X$ ; 该构式出现的总频数  $M$ ; 所有构式出现的总频数  $W$ 。本研究中, 词项在构式中出现的频数通过具体检索式获得;  $M$  值则通过检索式“[v \*] around”或“[v \*] round”在 COCA 语料库检索获得; 词项在所有构式中出现的总频数  $X$  等于该动词在语料库中出现的总频数; 所有构式出现的频数  $W$  则用所有动词出现的频数代替。b) 确定研究范围。经语料库检索发现, 出现于“V (a) round”构式的动词种类极其丰富, 根据研究目的, 本文各选取搭配频率最高的前 30 条进行分析。c) 搭配强度分析。计算各确定动词与构式关系的交叉表, 将表中的数据一一代入 SPSS 得出搭配强度, 根据 Fisher 值判定词素与构式之间是否具有一定的相互作用。d) 语义聚类分析。根据 SPSS 分析结果, 对 Fisher 值达到一定水平的动词进行语义归类和分析, 从而判定进入构式 V 槽动词的语义分类及构式的构式义。

### 三、结果和讨论

#### (一) 搭配频数和种类分析

通过 COCA 语料库检索发现, “V around”构式出现的频率远远大于“V round”构式, 且出现于“V around”构式 V 槽中动词的种类和频率都远远大于出现于“V round”中的动词。“V around”和“V round”构式吸引的高频动词也存在一定差异。

但由于每个动词出现的频数不一, 其与“V (a) round”构式搭配的频率也必然不同, 因此动词与构式共现频率高并不意味着构式搭配强度高。为进一步分析动词与构式的搭配强度, 本研究抽取与构式搭配频率最高的前 30 动词 (见表 2), 检索各词在“V (a) round”中出现的频率, 该词项在所有构式中出现的总频数  $X$ , 该构式出现的总频数  $M$ , 所有构式出现的总频数  $W$ , 并进行计算, 得到进行 Fisher 精确检验需要的四个数据  $a$ 、 $b$ 、 $c$ 、 $d$ , 以此计算共现频率前 30 的动词与“V around”构式或“V round”的搭配强度。以“revolve”为例, 经过检索和计算, 可以得到表 3 数据。

表 2 “V (a) round”构式高频动词搭配

V around 构式						V round 构式					
排序	词素	频率	排序	词素	频率	排序	词素	频率	排序	词素	频率
1	be	13344	16	drive	1734	1	be	632	16	wrap	36
2	look	12432	17	Stand	1502	2	go	461	17	run	34
3	turn	10446	18	spin	1429	3	come	270	18	grow	32
4	go	6547	19	glance	1400	4	turn	186	19	hang	30
5	walk	6493	20	build	1391	5	look	181	20	rally	28
6	get	4593	21	wander	1328	6	gather	152	21	whirl	28
7	run	4056	22	travel	1131	7	spin	90	22	open	27
8	come	4034	23	roll	1117	8	get	64	23	would	27
9	sit	3816	24	work	1060	9	hear	64	24	can	26
10	hang	3238	25	swirl	1004	10	swing	63	25	crowd	25
11	move	3204	26	fly	875	11	walk	61	26	help	24
12	wrap	2673	27	dance	844	12	have	52	27	glance	24
13	revolve	2469	28	float	839	13	work	49	28	wind	23
14	stick	1996	29	play	826	14	circle	44	29	twist	22
15	gather	1868	30	fool	822	15	will	44	30	make	21

表 3 “revolve”出现于“V around”的相关数据分析<sup>①</sup>

构式 词项	be	非 be	每行总数
[V around]	2469 *	163015 *	165484
非[V around]	658 *	96367237 *	96367895
每列总数	3127	96530252	96533379

将表 3 中带“\*”的四项数据代入 SPSS 中的交叉表进行 Fisher 精确检验, 得到 Fisher 值为 0.00

<sup>①</sup> 2469, 165484, 3127, 96533379 四项数据通过 COCA 语料库检索获得; 带“\*”的四项数据为 Fisher 精确检验所需, 其中, 163015, 658, 96367237 三项数据由计算获得。

( $p < 0.01$ ), 因此可以得出结论: 动词“revolve”与构式“V around”之间具有强烈的互动关系。依此类推, 计算上述两组数据共 60 个动词与相应构式的搭配强度, 得出以上两组动词与其对应构式之间均具有相互作用的结论。

为检验出现于“V round”构式的动词与“V around”是否也有一定关系, 对出现于“V round”的动词进行同样地操作, 发现与“V round”构式有互动关系的动词同时与“V around”构式有互动关系。

## (二) 基于数据结果的语义分析

由于以上动词均与构式具有强烈的互动关系, 可对这些能进入构式的典型动词进行语义分类, 发现进入“V (a) round”构式 V 槽的动词主要分为 2 类: a) 相对静止动词, 主要指不发生位移的动作, 如 stand、sit; b) 发生位置变换的动词, 这类动词又可以进一步分为 3 类: 表示由远及近的动作, 如 get、come、gather; 表示由近及远的动作, 如 look、go、hang、glance; 不具有方向, 表示在一定范围内发生的动作, 如 run、move、drive、circle、work 等; 进一步分析发现, 以上动词进入构式后凸显的部分也不相同, 结合实例<sup>①</sup>具体可以分为:

### 1. 对动作结果的突出

例 1: And again, basically, after tonight, you're not going to get around, whether by car, by train, by air.

例 2: It's charming and wonderful and helps business when you guys fool around and moan and groan in the middle of the night.

对以上两例进行分析发现, “get”进入构式后并没有凸显“get”的动作, 也不表示到处走动, 而是表示“摆脱、逃脱”某些令人不悦的事情, 因此突出的既不是动作本身, 也不是范围, 而是动作的结果; 同样例 2 中“fool around”突出的既不是“fool”的动作, 也不是“around”传达出的“范围、内容、路径”的含义。该构式此时传达的是“闲荡、游手好闲”的含义, 也是对事件结果的凸显。

### 2. 对方式、路线、范围的突出

例 3: People gathered round and started staring at us.

例 4: It whirls round as I push the door harder, as I am now on the way to the next place. Everything changes. Inside and outside.

例 3、例 4 均是对方式或路线的突出。例 3 中“gather round”中“gather”一词本身具有一定的方

向性, 表示由远及近的动作, “round”一词在这里给出一定的范围限制, 不是来自同一地点, 而是四面八方, 呈现出由面到点的运动方式; 例 4 中“whirls round”不仅强调门旋转的动作, 更强调了门转动的方式, 展现出门在一定范围内来回摆动的画面, 因此也是对方式的突出。

### 3. 对相邻、相关内容的突出

例 5: Myelin is a white, fatty tissue that glial cells wrap around the long tails of neurons, called axons.

例 6: The story told of Sindor and Haria in this text revolves around four dogs.

例 5 中, “wrap around”在句子中没有强调“wrap”的动作, 而是凸显了其后面的内容, 即“the long tails of neurons”, 因此是对动作相邻、相关内容的凸显; 同样例 6 中“revolves around”强调的也不是“revolves”动作本身, 而是紧跟其后的“four dogs”, 表明故事的内容是关于“four dogs”, 而不是其他。因此例 5、例 6 均是对动作相邻或相关内容的凸显, 而不是对动作本身或运动方式的凸显。

### 4. 对动作的持续性和连续性的突出

例 7: The Lentz family moved around, to Houston, Dallas, Philadelphia, Indianapolis and then back to Los Angeles.

例 8: Ledes said she saw a group of young girls wandering around in cheerleading outfits. (COCA)

例 7、例 8 凸显的是动作的持续性或重复性: “move around”由一系列“move”的动作组成, 是动作的连续和重复, 而“around”则用来限定移动的范围。因此, 该构式表达的是在一定活动范围内动作的不断重复, 凸显的是动作的反复性; “wandering around”也是对持续动作的强调, “wander around”凸显的是来回走动的过程, 是“走”这个动作的持续和反复。因此也是对动作持续性和连续性的凸显。

正如 Goldberg<sup>[1]152-179</sup>所说, 对构式义的理解就像是多义词, 在不同的语境下会呈现不同的意义。而之所以意义不同是因为构式凸显内容的不同。根据以上分析, “V (a) round”构式的构式义相应地可以分为四类: 表示动作的结果; 表示动作的方式、路线和范围; 表示动作相邻、相关内容; 表示动作发生时伴随的状态。

① 文中选取例句全部来自 COCA 语料库。



### (三) 构式搭配强度的分析对外语教学的意义

“V(a)round”中动词与构式搭配强度的分析以及进入该构式V槽的动词语义的分析对我国外语教学有如下启示:

a)以构式为单位进行语言教学。以单词或是词组为单位,遵循传统语法规则进行教学虽为我国的外语教学做出了巨大的贡献,但仍存在一定不足。构式语法打破词汇和语法的界限,以构式为语言单位,将词汇和结构融合在一起,可弥补传统语法无法解释的一些语言现象且可避免学生常见的一些搭配错误。

b)借助大型英语语料库进行构式及其语义的学习。大型英语语料库如COCA、BNC等语料库等不仅容量大,而且覆盖范围广,语料来源包括小说、口语、报刊等,不仅可以为我们提供大量的语言实例,而且借助语料库进行数据分析,根据数据统计结果分析构式义的方法也更加全面、科学、有效。

### 四、结语

本文借助COCA语料库对出现于“V(a)round”构式的动词进行检索,使用Stefanowitsch和Gries的构式搭配分析法分析动词与构式的搭配强度,并对与“V(a)round”构式具有强烈互动关系的动词进行语义聚类分析,结果发现出现于“V around”构式的动词虽在频率和种类上远远大于出现于“V round”构式的动词,但与“V round”构式具有显著互动关系的动词同样与“V around”构式具有强烈互动关系。此外,能够获得“V(a)round”构式允准进入构式的动词可以是相对无位置变化的动词,也可以是表示由远及近、由近及远或是运动轨迹不规律的动词。而进入“V(a)round”构式的动词不同会造成构式凸显的不同,从而造成该构式在不同情况下发生表示动作结果、方式、路线和范围,动作相邻、相关内容,或是伴随状态的变化。借助Stefanowitsch和Gries的构式搭配分析法,对“V(a)round”构式的搭配强度及其语义进行分析,一方面为构式学习提供新的视角,另一方面也有助于该构式搭配分析法的进一步发展和运用,为更多的构式及构式义研究提供一定的参考价值。

### 参考文献:

[1] GOLDBERG A. E. *Constructions: A Construction Grammar Approach to Argument Structure* [M]. Chicago: The University of Chicago Press, 1995.

[2] 王仁强,陈和敏.基于语料库的动词与构式关系研究:以sneeze及物动词用法的规约化为例[J].外语教学与研究,2014,46(1):19-31.

[3] 张建理,骆蓉.构式、整合与语言集约化:对Away构式的调查[J].外国语(上海外国语大学学报),2014,37(6):34-42.

[4] 王寅.“新被字构式”的词汇压制解析:对“被自愿”一类新表达的认知构式语法研究[J].外国语(上海外国语大学学报),2011,34(3):13-20.

[5] 魏晓龙,高原.作为评价手段的构式:X+BE+being+ADJ[J].解放军外国语学院学报,2016,39(1):62-71.

[6] 方强,张萍.构式对词项主观性的压制研究[J].现代外语,2016,39(1):22-30.

[7] 许红菊.“(S)有的是NP”句的语义生成原理和习语化历程:基于语法构式和修辞构式的分类理论[J].语言研究,2016,36(4):28-34.

[8] 杨江峰.基于语料库的原型理论视域下汉语迂回致使结构研究[J].外语教学,2016,37(2):7-11.

[9] 袁红梅,梁婧玉.“被+X”构式义的概念整合分析[J].外语研究,2016,33(1):33-39.

[10] 席建国,王文斌. Have a N<sub>1</sub>(Dev)结构之语法转喻及构式化特征考察[J].外语研究,2016,33(1):18-23.

[11] 杨永龙.从“形+数量”到“数量+形”:汉语空间量构式的历时变化[J].中国语文,2011,(6):499-513.

[12] 张建理,徐银.构式语法对汉语“形容词+宾语”结构的研讨[J].外国语(上海外国语大学学报),2011,34(6):11-18.

[13] 张克定.汉语处所主语形容词谓语句式的认知机制和聚焦过程[J].现代外语,2016,39(1):11-21.

[14] 吕建军.“王冕死了父亲”的构式归属:兼议汉语存现构式的范畴化[J].语言教学与研究,2013(5):75-83.

[15] 甄珍.现代汉语主观极量构式“要多A有多A”研究[J].汉语学习,2015(1):57-66.

[16] 罗思明,查如荣,江晶晶.英汉“NA表量构式”中形容词制约的语料库与类型学研究[J].外语研究,2014,31(3):31-37.

[17] 罗思明,王文斌,洪明.英汉结果构式R<sub>1</sub>(AP)制约的语料库与类型学研究[J].外语教学与研究,2010,42(4):268-274.

[18] 陈佳.基于语料库的“COME/GO+形容词”构式搭配关联强度与构式范畴化关系研究[J].解放军外国语学院学报,2015,38(3):23-30.

[19] STEFANOWITSCH A, GRIESS T. Collostructions: Investigating the interaction between words and constructions [J]. International Journal of Corpus

- Linguistics, 2003, 8(2): 209-243.
- [20] CHURCH K W, GALE W, HANKS P et al. Using statistics in lexical analysis [C]// ZERNIK U (ed.). Lexical Acquisition: Exploiting Online Resources to Build up a Lexicon. Hillsdale. NJ: Lawrence Erlbaum, 1991: 115-164.
- [21] DUNNING T. Accurate methods for the statistics of surprise and coincidence [J]. Computational Linguistics, 1993, 19(1): 61-74.
- [22] 杨晶, 王勇. into-致使构式中互为变化的共现词素: 国外对 into-致使构式研究综述[J]. 外语学刊, 2010(2): 70-75.
- [23] 胡健, 张佳易. 认知语言学与语料库语言学的结合: 构式搭配分析法[J]. 外国语(上海外国语大学学报), 2012, 35(4): 61-69.
- [24] 田臻, 吴凤明, 曹娟. 英汉存在构式与动词语义关联的实证对比研究[J]. 外语教学与研究, 2015, 47(6): 826-837.

## Study on Collostructional Strength of “V (a) round” Construction and Semantic Clustering of “V” Verb Based on Corpus

YANG Shasha

(School of Foreign Languages, Guangxi University, Nanning 530004, China)

**Abstract:** The verbs of “V (a) round” construction were searched in COCA corpus, and Fisher was employed to precisely check the collostructional strength. The findings show that the verbs occurring in the “V (a) round” construction mainly include two categories: static verbs and transitional verbs; the emphasis of construction varies from verb type to verb type, which is mainly on consequence of actions, or way, path and range, or neighboring and relevant content, or action durability and continuity; accordingly, the construction also varies, meaning consequence of actions, or way, path and range of actions, or neighboring and relevant content of actions, or state accompanying the action occurrence. The analysis of collostructional strength of “V (a) round” construction and semantic thereof based on corpus cannot only helps us to have a better understanding of this construction, but also is a further application of the collostructional analysis Method. It is also enlightening for foreign language teaching for Chinese universities.

**Key words:** collostructional strength; “V (a) round” construction; constructional meaning; collostructional analysis method

(责任编辑: 任中峰)