

基于语料库的词汇深度分析

——以致使结构“Get Treated”为例

范若澜

浙江工商大学外国语学院, 浙江 杭州

收稿日期: 2023年10月9日; 录用日期: 2023年11月13日; 发布日期: 2023年11月23日

摘要

语料库工具提供了大量真实的词汇搭配实例, 使得对词汇特点的深入观察和分析成为可能, 为词汇研究提供了新的视角和途径。本研究基于美国当代英语语料库(COCA), 以Sinclair的扩展意义单位模型为理论框架, 从类联接、搭配、语义倾向和语义韵等四个维度对“get treated”这一致使结构进行分析。研究揭示了“get treated”在语境中的具体使用方式、搭配、语义特点等, 促进了对其词汇知识的全面掌握, 并为词汇深度学习提供有益的启示和参考。

关键词

语料库, 类联接, 搭配, 语义倾向, 语义韵

A Corpus-Based Study of Vocabulary Depth

—Taking the Causative Structure “Get Treated” as an Example

Ruolan Fan

School of Foreign Language, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: Oct. 9th, 2023; accepted: Nov. 13th, 2023; published: Nov. 23rd, 2023

Abstract

Corpus tools provide a large amount of authentic instances of collocations, enabling in-depth observation and analysis of the features of vocabularies, as well as offering new perspectives and approaches for research in the related field. Based on the Corpus of Contemporary American English (COCA), this study takes Sinclair's Extended Unit of Meaning as the theoretical foundation, and analyzes the causative structure “get treated” from four dimensions: colligation, collocation, semantic preference, and semantic prosody. This study sheds light on the specific usage, collocations,

and semantic characteristics of “get treated” in the context, contributing to a comprehensive understanding of its vocabulary knowledge, as well as providing valuable implications for vocabulary learning.

Keywords

Corpus, Colligation, Collocation, Semantic Preference, Semantic Prosody

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

语料库的兴起为语言学领域的实证研究提供了工具，使得对大量真实存在的语言的分析成为可能。随着语料库语言学的发展，学界把单词(word)作为最小意义承载单位的传统观念被打破。Sinclair 基于语料库的词汇共现研究发现提出了短语理论，将意义的承载单位扩展到了词组，并在此基础上提出了扩展意义单位模型[1] [2]。在该模型的指导下，研究者以真实语料为线索，考察词汇的搭配、类联接、语义倾向和语义韵，获得词汇的深度知识。扩展意义单位模型的提出具有理论和实践意义：不仅为研究者提供了意义研究的新视角，将以往的研究热点从词汇广度转移到词汇深度，对语篇分析、语言习得等领域的研究产生了深远影响；其研究结果还被应用于词典编纂、词汇教学等实践活动。鉴于此，本文尝试借助语料库工具，以扩展意义单位模型为理论框架，分析致使结构“get treated”的搭配、类联接、语义倾向和语义韵特征，获得该结构的词汇深度知识，为其今后在具体语境中的正确使用提供语料库数据支撑，并为词汇深度学习提供有益的启示和参考。

2. 文献概述

Sinclair 提出的扩展意义单位模型为词汇和短语意义研究提供了系统的理论框架，从搭配、类联接、语义倾向和语义韵四个维度来了解词汇和短语的属性和特征[2]。搭配(collocation)指的是“两个多更多的词在文本有限跨度内的共现”，强调一组词作为整体的意义而非单个词的意义。类联接(colligation)与搭配类似，都关注词语的共现关系，但前者更为抽象，是节点词习惯呈现的一种语法结构，反映的是语法范畴间的共现关系，包括介词、名词、动词、副词等[3]。而语义倾向(semantic preference)是与节点词共现的搭配词的语义特征，也是对搭配词在语义上的限制。根据语义的区别，这些搭配词往往可以分为不同的语域，相同语域内的搭配词共享相同的语义成分。作为语料库语言学发现中最重要的语言运作机制之一[4]，语义韵(semantic prosody)具有强制性，决定着与节点词共现的语法范畴、搭配词及其语义特征，在词组的意义构建中发挥重要作用[2]。语义韵可以简单地理解为一种体现说话人或作者态度或情感的意义，具有语用功能特征[5]。Louw 将语义韵大致分为积极、中性和消极三种类别[5]，而后来的语言学家们在研究中打破了这种局限，将语义韵作了更为具体和细致的概括[6]。因此，扩展意义单位模型将词汇、语法、语义和语用四个要素结合在一起，为获取较为全面的词汇深度知识提供了科学的理论支持。

当前，基于扩展意义单位的词汇深度知识研究已成为国内语言学界的热门话题，尤其是语义韵这一概念，已经成为了语料库语言学研究中的重要对象[7]。例如，王家义，李德凤利用扩展意义单位对英语词汇“commit”进行了深入分析[8]。同样，扩展意义单位模型也能应用于汉语语料的分析[9] [10] [11]。

此外,基于汉英语料库的词汇对比分析也是热门研究方向,常用于揭示中国学生在词汇学习过程中对搭配、语法、使用等情况的掌握程度,为词汇教学实践带来新的启示[12][13][14]。

3. 研究设计

本研究使用的是美国当代英语语料库(The Corpus of Contemporary American English, 简称 COCA)。COCA 语料库由 10 亿词的文本构成,涵盖 1990 至 2019 年间美国的口语及书面文本,其语料均衡分布在小说、杂志、报刊、学术期刊、口语、博客、网页、影视剧等多种体裁中。因此,该语料库提供了丰富的英语母语者在生活中的语言使用实例,推动了语料库语言学的发展和研究。

利用基于数据的方法,本研究将致使结构“get treated”作为节点词在 COCA 语料库中进行搜索,将搜索结果进行收集和整理得到一个由 411 条索引行构成的微型语料库。再以 WordSmith6 为工具观察索引行,结合扩展意义单位模型对“get treated”的语域分布、搭配、类联接、语义倾向和语义韵等特点进行分析。

4. 分析与讨论

4.1. 语域分布

致使结构“get treated”在 COCA 语料库中的出现频率为 411 次,经手工剔除重复出现的 8 条结果后,总计出现频率为 403 次。“get treated”的分布主要集中于博客(27%)、口语(23%)、网页(16%)和影视剧(14%)中,而在报刊(8%)、杂志(7%)、小说(6%)和学术期刊(2%)中的分布较少。可见,“get treated”具有明显的口语特征,不常用于书面语。

4.2. 搭配与类联接

据统计显示,致使结构“get treated”在 COCA 语料库中出现 403 次。本研究主要观察“get treated”短语两边 L3 至 R3 位置上的词(详见表 1)。在 R1 的位置上,介词(45%)与致使结构“get treated”有很强的共现倾向,包括“like”、“for”、“as”、“with”等,构成短语“get treated like/for/as/with sth.”,意为“被当作……来看待”或“得到……的对待”,而当介词为“for”时,则意为“接受(对某种疾病)的治疗”。有 22%的副词来修饰短语“get treated”,构成“get treated + adv.”的结构,表示“得到……的对待”,包括“differently”、“better”、“any”、“fairly”、“well”等。当 R1 的位置出现“and”(3%)、“or”(1%)、“that”(1%)等连词时,其后常跟随句子,通常不直接修饰“get treated”或与其搭配。此外,在 14.9%的情况下“get treated”被置于句尾,右边不跟任何搭配词。在 R2 的位置上,定冠词“the”(6%),不定冠词“a”(9%)和“an”(2%)的出现频率较高,主要用于修饰 R3 位置上的名词。形容词“same”(2%)与 R3 上的“way”构成“get treated the same way”的结构,表示“得到同样的对待”。而介词“than”(2%)与 R1 上的副词构成诸如“get treated better than”等结构,同样修饰“得到了……的待遇”。17%的名词与 R1 上的介词构成如“get treated like sth.”的短语,其中包括“crap”、“shit”等词以及表示疾病的名词。有 12%的副词,如“better”、“well”、“differently”、“worse”等,与 R1 上“much”、“so”、“any”等表示强调的程度副词组合在一起构成“get treated much/so/any + adv.”结构。而 R2 上的形容词(6%)多用于修饰 R3 上的名词,包括“trivial”、“similar”、“different”等。当 R3 的位置上出现“queen”、“asshole”、“jerk”等名词(17%)时,与前面的介词构成完整的“get treated like/with/as/for/to sth.”表达式。同时,形容词(7%)与副词(1%)也通常与 R4 上的名词或 R2 上的副词搭配修饰“get treated + prep”结构。介词“as”(2%)通常以“as...as”的形式出现,而“by”(2%)则构成“get treated like/with/as/for/to sth. by sb.”的组合,表示“被某人如何对待”。剩余情况下,其他词汇通常不与“get treated”结构搭配或直

接修饰该结构。

L1 位置上出现频率最高的词为“to”(14%),与后项构成不定式结构“to get treated”,而前项往往是动词(60%)、形容词(24%)和情态动词(16%)。“you”、“they”、“I”、“we”等代词(39%)则表示“对象”,构成“sb. get treated+ prep./adv.”的表达。此外,助动词与情态动词占了19%,如“do”、“can”、“could”、“will”、“should”等,而包括“don’t”、“not”在内的否定表达则占了12%。在L2的位置上,有51%的搭配词属于代词和名词,包括“you”、“they”、“I”、“we”等,修饰“get treated”的对象。“that”、“and”、“how”、“where”、“but”等连词(11%)用于衔接后面的短句。剩余的搭配词还包括助动词和情态动词(8%)、动词(8%),修饰主体的意愿或动作,构成“aux./modal verb/do (to) get treated”的结构。L3位置的词多为功能词,且通常不直接修饰“get treated”结构。因此,致使结构“get treated”主要与L3-R3位置上的词搭配,且与右边词汇搭配的趋向更明显。该短语与其搭配词主要构成“sb. get treated + prep.(like/for/as/with/to) + sth.”和“sb. get treated+ adv.”结构,表示“某人受到……的对待”。

Table 1. Collocations of “get treated” in COCA

表 1. “get treated” 在 COCA 语料库中的搭配情况

L3	L2	L1	center	R1	R2	R3
				like (84)		
	you (17)			for (26)		
the (20)	they (12)	to (56)		as (21)	a (36)	
and (13)	that (12)	you (42)		with (13)	the (24)	
that (12)	do (11)	they (33)		the (12)	same (9)	and (10)
you (8)	the (10)	don't (23)		differently	than (8)	as (9)
I (8)	I (10)	and (21)		(12)	an (8)	you (8)
a (8)	we (9)	I (18)		better (12)	this (7)	the (8)
of (7)	hospital (9)	not (13)	get treated	and (12)	crap (7)	by (8)
like (7)	and (9)	we (11)		in (11)	way (6)	way (7)
if (7)	going (8)	will (10)		to (10)	one (6)	they (6)
to (5)	people (7)	women (9)		by (9)	get (6)	I (6)
they (5)	did (6)	still (7)		you (7)	and (6)	care (5)
how (5)	way (5)	people (6)		at (7)	that (5)	
	how (5)	men (6)		any (7)	shit (5)	
	he (5)			fairly (6)		
				well (5)		

(注: 搭配词出现的最小频率为5)。

4.3. 语义倾向与语义韵

通过观察 403 条索引行, 本研究发现致使结构“get treated”搭配词的语义特征可大致分为三类: “体现贬义、消极意味”、“表示尊重、友好的态度”以及“传递客观信息的表达”(详见表 2)。

Table 2. Semantic preference and semantic prosody of “get treated”

表 2. “get treated” 搭配词的语义特征

语义特征	“get treated” 的搭配词
体现贬义、消极意味	get treated like scum ; get treated like a piece of shit ; get treated like trash ; get treated like punk ; get treated like a handicapped person ; get treated as trivial ; get treated like second-class citizens ; get treated as bad as the Hygienist ; get treated badly ; get treated horribly ; get treated like garbage ; get treated more violently ...

Continued

表示尊重、友好的态度	get treated like celebrities ; get treated like the Queen of England ; get treated equally ; get treated fairly ; get treated with respect ; get treated like royalty ; get treated properly ; get treated well ...
传递客观信息的表达	get treated at the Emergency room; get treated by a sport medicine specialist; get treated by the first doctor; get treated for breast cancer; get treated for bacterial infection; get treated for injuries; get treated in time...

据表 2 显示, 当“get treated”后接介词或副词时, 其搭配的词汇可根据语义特征分为“体现贬义、消极意味”、“表示尊重、友好的态度”和“传递客观信息”三类。“get treated like/as/with”结构表示“像……一样被对待”时, 后面所接的名词多为“败类”、“混蛋”、“垃圾”、“流氓”、“尘埃”、“二等公民”等带有贬义色彩的蔑称。虽然“get treated”偶尔会搭配“女王”、“名人”、“尊敬”、“王室成员”等表现尊敬、友好态度的名词, 但这时“get treated”的前项通常是表现否定意义的词汇, 如“don't”、“can't”、“will not”、“shouldn't”等, 或是表现主体个人意愿的表达, 如“hoping to get treated to a little primping and pampering”。而当“get treated”后接介词“for”和时, 其语义发生了变化, 表示“接受(对某种疾病)的治疗”而非“受到……的对待”。因此, 其后经常搭配疾病类名词, 用于描述客观事实。在另一个高频结构“get treated + adv.”中, 其搭配的副词可分为带消极色彩(“糟糕地”、“粗暴地”、“严厉地”等)和带积极色彩(“平等地”、“公平地”、“良好的”等)两类。同样, 在表现尊重、友好态度的副词前通常会出现否定词来抵消其积极意义。

因此, 致使结构“get treated”的语义韵特征可分为两类。当语义为“被……对待/看待”时, 其搭配词体现贬义、消极意味或是带有否定尊重、友好态度意味的表达。受这些搭配词的影响, “get treated”也呈现出消极的语义韵特征, 表示“被轻蔑地对待/看待”。而当语义表示“接受治疗”时, 其搭配词不具有特定语义色彩, 促使“get treated”的语义韵呈现出中性的特征。

5. 小结

本研究在 COCA 语料库中检索致使结构“get treated”, 通过语料库分析软件 WordSmith6 对该结构的真实使用情况进行观察, 并以扩展意义单位模型为理论框架对“get treated”的搭配、类联接、语义倾向和语义韵的分析。研究发现“get treated”具有显著的口语特征, 不常用于书面语。在真实的语言使用环境中, 该结构常以“sb. get treated + prep. + sth.”和“sb. get treated+ adv.”的形式出现。当“get treated”意为“被……对待”时, 其搭配的词汇通常具有贬义意味和消极语义特征, 使得整个结构呈现出消极的语义韵特征, 表示“被轻蔑地对待/看待”; 而当“get treated”表示“接受治疗”时, 搭配词往往用于传递客观的信息, 不具有特别的语义特征, 促使该结构的语义韵呈现出中性的特征。因此, 本研究基于语料库驱动的方法, 揭示了致使结构“get treated”的特征及词汇深度知识, 并为词汇深度学习提供有益的启示和参考。

参考文献

- [1] Sinclair, J. (1996) The Search for Units of Meaning. *Textus Online Only*, 11, 75-106.
- [2] Sinclair, J. (2004) *Trust the Text: Language, Corpus and Discourse*. Routledge, London. <https://doi.org/10.4324/9780203594070>
- [3] 濮建忠. John Sinclair 的短语理论与意义研究[J]. 当代外语研究, 2021(6): 60-76.
- [4] 卫乃兴. 基于语料库学生英语中的语义韵对比研究[J]. 外语学刊, 2006(5): 50-54.

- [5] Louw, B. (1993) Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In: Baker, M., Francis, G. and Tognini-Bonelli, E., Eds., *Text and Technology: In Honour of John Sinclair*, John Benjamins, Amsterdam, 157-715. <https://doi.org/10.1075/z.64.11lou>
- [6] 卫乃兴. 共选理论与语料库驱动的短语单位研究[J]. 解放军外国语学院学报, 2012, 35(1): 1-6, 74.
- [7] 李兰芳. 卫乃兴. 语义韵属性再探究——基于汉语语料库的分析[J]. 广西民族大学学报, 2015, 37(4): 171-176.
- [8] 王家义. 李德凤. 基于 BNCweb(CQP-Edition)的英语词汇深度分析——以“commit”为例[J]. 外国语文, 2018, 34(6): 102-110.
- [9] 田宏梅. 利用汉语语料库研究词语搭配——以“有点”为例[J]. 暨南大学华文学院学报, 2006(3): 67-73.
- [10] 吴海霞. 基于语料库的汉语量词语义韵考察——以“群”“帮”“伙”为例[J]. 湖北科技学院学报, 2018, 38(2): 59-65, 74.
- [11] 杨蓉. 基于扩展意义单位理论的多维译义探讨——以“教育”为例[J]. 外国语言与文化, 2018, 2(3): 135-145.
- [12] 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, 35(6): 438-444.
- [13] 王海华, 王同顺. CAUSE 语义韵的对比研究[J]. 现代外语, 2005, 28(3): 297-307.
- [14] 张继东, 刘萍. 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[J]. 外语研究, 2006(5): 19-22.