

基于COCA语料库的近义词辨析

——以Vague、Obscure和Ambiguous为例

段彩玉

曲阜师范大学, 翻译学院, 山东 日照
Email: 582588660@qq.com

收稿日期: 2020年11月15日; 录用日期: 2020年11月30日; 发布日期: 2020年12月7日

摘 要

英语词汇中的近义词辨析一直是英语学习中的重中之重, 而学生们通常采用死板的记忆方式而忽略了其本身的涵义, 往往会导致不会用甚至错用的现象。近年来语料库的出现及应用为英语学习者们带来了很大的便利, 尤其是对英语中近义词的辨析与学习。本文基于美国当代英语语料库, 选取vague、obscure和ambiguous三个具有“模糊的”意义的形容词为研究对象, 多方面多维度地对其词频、搭配及语义韵特征等进行对比性分析, 希望为英语学习提供新的视角与帮助, 阐释在教学中如何运用语料库能更准确地把握近义词的区别, 帮助学生提高英语词汇学习的兴趣与水平。

关键词

COCA语料库, 同义词, 语域, 语义韵

Synonym Discrimination Based on COCA Corpus

—Taking Vague, Obscure, and Ambiguous as Examples

Caiyu Duan

School of Translation Studies, Qufu Normal University, Rizhao Shandong
Email: 582588660@qq.com

Received: Nov. 15th, 2020; accepted: Nov. 30th, 2020; published: Dec. 7th, 2020

Abstract

Synonyms discrimination in English vocabulary has always been the top priority in English learn-

ing, and students usually adopt a rigid way of memorizing but ignore its own meaning, which often leads to the phenomenon of unable to use or even wrong use. In recent years, the emergence and application of corpus have brought great convenience to English learners, especially the discrimination and learning of synonyms in English. Based on the Corpus of Contemporary American English (COCA), this paper selects vague, obscure and ambiguous as the object of study, which have the meaning of “模糊的”, and makes a comparative analysis of their word frequency, collocation and semantic prosody characteristics in many aspects, hoping to provide new perspectives and help for English learning and explain how to use corpus in teaching to grasp the difference of synonyms more accurately so as to help students improve their interest and level in English vocabulary learning.

Keywords

COCA Corpus, Synonyms, Register, Semantic Prosody

Copyright © 2020 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

在英语词汇学习中, 我们不难发现有许多具有相似意义的动词需要根据其用法和搭配进行辨析和区分, 所以以往的辨析研究主要集中在动词, 而由于英语形容词意义的多样性, 我们往往忽略了英语形容词的辨析而导致了許多错用现象。举一个比较常见的例子, 在许多公共场合中, 表示“保持安静”的公示语中多采用“quiet”而不用“silent”。虽然形容词在句子中仅承担着修饰成分, 但其本身的性质特征及状态与其后所形容的名词或语境也需要搭配, 即便是微小的差异也可能会让人产生误解。

既然我们知道了英语词汇辨析的重要性, 那么我们就需要掌握英语词汇辨析的方法。以往教师们往往根据参考书和词典的解释进行讲解, 结果发现效果并不理想, 学生还是会很难区分, 没有正确的区分就可能没法做到理解和运用。语料库的兴起和应用为英语词汇的辨析提供了另一种可能, 本文将选取 vague、obscure 和 ambiguous 三个形容词作为研究对象, 分析其在语料库中不同语域的词频、搭配、语义韵等三个方面的语言特征, 从而比较这组近义词的用法差异, 阐述语料库在英语教学中的运用, 给英语教学和学习带来一些启发, 使学生对词汇的表达更贴近母语者的表达方式。

2. 研究工具及方法

目前, 随着网络技术的快速发展, 各种大型的语料库孕育而生, 本文选用了世界最大的英语平衡语料库——美国当代英语语料库(Corpus of Contemporary American English, 简称 COCA)为研究工具。COCA 语料库收集了口语, 小说, 杂志, 报纸以及学术期刊等五种不同文体, 词汇量较大, 而且每年会新增 2000 万的新词, 时效性强。

主要研究方法为统计 vague、obscure 和 ambiguous 在不同语域中的分布情况, 显著搭配词以及观察其搭配特征, 总结内在的类联结和语义韵(张继东, 刘萍, 2005 [1])。

3. 基于 COCA 语料库的对比分析

通过 COCA 语料库中的词频列表检索、搭配检索、以及对比检索, 对目标词的词频、语域、搭配关系和语义韵等方面进行观察和分析, 总结 vague、obscure 和 ambiguous 三个形容词的词频对比、不同语

域的词频分布、搭配词的显著差异。

3.1. 词频对比

Table 1. Statistics of word frequency distribution of vague, obscure and ambiguous in COCA Corpus

表 1. Vague、Obscure 和 Ambiguous 在 COCA 语料库中的词频分布统计

词项	总频数	形容词总频数
vague	6630	6630
obscure	5234	3504
ambiguous	3773	3773

当我们从词典上查这三个词的词义，可以发现 vague 解释为模糊的、含糊的、不明确的；obscure 解释为昏暗的、朦胧的、模糊的还有不清楚的、隐蔽的等意思；而 ambiguous 有模糊不清的、模棱两可的、不明确的等意思。相对于 vague 和 ambiguous 两个词，obscure 的含义更加丰富，并且它可以做名词和动词使用。

从表 1 可以看出，这三个近义词按其出现频率由高到低排序为：vague、obscure、ambiguous，而当只看形容词的总频数的时候，vague 的总频数几乎是 obscure 和 ambiguous 的两倍，obscure 的频数明显降低，说明说明当人们想要表达“模糊的”的语义时更加倾向于使用 vague 和 ambiguous，但其实 ambiguous 和 obscure 的频数差距不大。而 obscure 当名词和动词使用的时候也可以表示“使...模糊不清，掩盖；隐藏；使难理解”或“某种模糊的或不清楚的东西”。因为本研究主要研究三个形容词的辨析，因此之后的检索都只检索 obscure 的形容词词性的用法。

3.2. 不同语域中的词频分布

杨慧中(2002 [2])认为语域是指人们由于某种交际的需要，而在不同场合或情景下，产出不同语体风格、语格调等的语言变体。尽管近义词的意思非常相近，但是它们在不同的语域中出现的频率会有不同并且呈现不同特征，即便在同一语域中也会表现出优先选择的差异。因此词汇在语域中的分布状态对于语言学习者来说至关重要。下面将对 vague、obscure 和 ambiguous 在不同语域中的频率进行统计。

Table 2. The total frequency of vague, obscure and ambiguous in COCA

表 2. Vague、Obscure 和 Ambiguous 在 COCA 五种语域中出现的总频数

语域	vague		obscure		ambiguous	
	词频	频率	词频	频率	词频	频率
		(permil)		(permil)		(permil)
口语	724	6.20	312	2.67	303	2.60
小说	2231	19.95	745	6.66	308	2.75
杂志	1247	10.63	969	8.26	545	4.64
报纸	910	8.05	784	6.94	364	3.22
学术	1518	13.63	694	6.23	2253	20.22

表 2 显示，vague 无论是在口语、小说、杂志和新闻中的使用频率都高于另外两词，说明在具体的语境中，vague 的使用频率更高。就 vague 而言，在小说文本中使用最多，高达 2231 次，小说的用词比较

形象生动,这说明 *vague* 的用词范围广。并且在学术文本中的使用频率次之,在口语中最少。与 *vague* 类似, *ambiguous* 一词也更常见于学术期刊中,只是远比 *vague* 的使用频数更多,而它在其他四个文本中的使用频率都不高,说明人们在学术文本中更喜欢选用 *ambiguous*, *ambiguous* 本身的专业性最强,更适用于专业文本。而相较于前两个词来说, *obscure* 的使用更加广泛,它在小说、杂志、报纸和学术期刊四个语域中的使用频率比较均衡,在口语语域中的使用频率最低,杂志领域使用频率最高,说明 *obscure* 更多用于文本写作而不是口语化的语域中。

3.3. 搭配词的显著差异

根据单词的不同搭配现象,我们可以观察出其搭配习惯,从而掌握词汇的用法。对于一个词来说其同一搭配词出现的频率越高,则表明其与目标词的关联性越强。MI 值(Mutual Information Score)是指将两个词同时出现的可能性和单独出现的可能性相比较,如果 MI 值为 0,那就表示这两个词基本不同时出现;如果 MI 值大于 3 或更高,就意味着这两个词经常互相搭配一起出现。本文检测了 *vague*、*obscure* 和 *ambiguous* 的搭配词以及搭配词出现的频数及 MI 值,检索时以该单词的前 5 和后 5 个单词为跨距(-5/+5)。此处仅取出现频数为前 15 的搭配词以作示例,按 MI 值从大到小排列。

Table 3. The first 15 collocation of *vague*, *obscure* and *ambiguous* in COCA

表 3. *Vague*、*Obscure* 和 *Ambiguous* 在 COCA 中的频数为前 15 的搭配词

词项	<i>vague</i>	<i>obscure</i>	<i>ambiguous</i>
搭配词	recollection/deliberately/notion/ memories/memory/symptoms/ answers/somewhat/sense/feeling/ concept/details/term/idea/terms/	origins/tends/relatively/provision/clouds/ somewhat/reference/remain/previously/ remained/facts/remains/reasons/meaning/ named	morally/statute/situations/somewhat/ status/term/meaning/remains/complex/ language/position/situation/nature/ relationship/results

将表 3 中每个词的频数为前 15 的搭配词进行分类,可观察到三个词都趋向于和名词搭配,和副词搭配的时候很少。与 *vague* 搭配的副词为 *deliberately*、*somewhat*; 与 *obscure* 搭配的副词有 *relatively*、*somewhat*、*previously*; 与 *ambiguous* 搭配的副词有 *morally*、*somewhat*。不难发现,三个词共有的一个搭配副词是 *somewhat*。

数据表明,节点词 *vague* 在 COCA 中的显著搭配词大致可分为 6 类: 1. 概念类: *notion*、*concept*; 2. 回忆类: *recollection*、*memories*、*memory*; 3. 症状类: *symptoms*; 4. 感觉类: *sense*、*feeling*; 5. 细节类: *details*、*terms*、*term*; 6. 想法类: *idea*、*answers*。需要指出的是,在众多搭配词中,表示概念类、回忆类和感觉类的搭配词总体使用比例比较高,并且词汇丰富程度或多样性高。通过语境确认,发现 *vague* 更多用于形容模糊的概念、记忆和感觉方面。节点词 *ambiguous* 在 COCA 中的显著搭配词大致可分为 5 类: 1. 环境、处境类: *situation*、*situations*、*nature*、*position*、*status*; 2. 概念类: *meaning*; 3. 语言类: *language*; 4. 关系类: *relationship*; 5. 结果类: *result*。可以表明 *ambiguous* 多与表示环境、处境及位置的词语搭配使用。节点词 *obscure* 的搭配比较特殊,它较多与动词搭配,例如: *tends*、*remain* 和 *named*。除此之外其名词搭配词在 COCA 中的显著搭配词大致可分为 5 类: 1. 起源类: *origins*; 2. 规定类: *provision*、*reference*; 3. 事实类: *facts*; 4. 原因类: *reasons*; 5. 概念类: *meaning*。纵观 *obscure* 名词性搭配词的特点,可以看出节点词 *obscure* 的名词搭配词多为抽象名词。综合以上数据可以看出三个词共有的搭配词为表示概念类的词,例如: *meaning*。

4. *Vague*、*Obscure* 和 *Ambiguous* 的语义韵对比研究

语义韵是语料库语言学迄今为止发现的重要语言运作机制之一,它是指词项与其搭配词在文本中反

复共现时, 呈现出相关的语义特点, 在语境内产生特定的语义氛围(卫乃兴, 2006 [3]), 代表词汇的一种搭配倾向性, 它们习惯性地吸引具有相同或相似语义特点的词项, 与之构成搭配关系。Louw (2000 [4]) 认为语义韵的主要功能是表达说话者或作家的态度或评价, 他指出要关注一个节点词或短语的搭配, 强调语义韵的语用功能。Stubbs (1995 [5]) 从功能的角度将语义韵划分为积极、中性和消极三类。研究近义词的语义韵能够更好的帮助二语学习者辨析近义词的异同, 从而更好的掌握词汇之间差异。

本文选取 vague、obscure 和 ambiguous 修饰后置名词的搭配词, 跨距设为 0/+5, 搜索这三个词右侧五词以内的显著搭配词(MI 值大于 3), 且按出现频数排序, 选择其中前 100 个显著搭配词进行分析, 突显这三个词“模糊的”的对象, 通过表 4 观察它们的语义韵特征。

Table 4. The distribution of the first 100 significant collocation of vague, obscure and ambiguous in positive, negative and neutral semantic prosody

表 4. Vague、Obscure 和 Ambiguous 的前 100 个显著搭配词在积极、消极和中性三个语义韵方面的分布表

	积极	消极	中性	总数
Vague	1	35	64	100
Obscure	2	7	91	100
Ambiguous	1	5	94	100

从前 100 个显著搭配词中, 发现 vague 的显著搭配词如 threats、unease、rumors、discomfort、warning、suspicion、anxiety、uneasiness、dissatisfaction、allegation、complaints、dread、apology 等词, 都是一些带有焦虑、恐惧、威胁、怀疑等意义的词语, 这些词语都带有消极的语义氛围。而其与中性的语义词搭配的情况多为表示回忆、概念、感觉类的词语, 例如: sense、feeling、impression、memory、recollection、notion、concept 等词。统计发现 vague 与消极词汇的搭配有 35 个, 与中性词汇的搭配有 64 个, 而与积极词汇搭配的语义韵只有 1 个, 因而 vague 具有明显的消极语义韵。

Ambiguous 的前 100 个显著搭配词的种类比较多, 例如环境类: circumstances、situation、status、position 等; 方向类: orientation、signals 等; 语言类: language、phoneme、phrase、message、clause、passage 等; 金钱类: legacy、figures 等, 这些词没有明显的情感偏向, 都是中性的词汇, 其频数达 94 个词。与 ambiguous 搭配的词还有带有消极语义韵的词汇如: provocation、statement、ambiguity、stimuli、syndrome 等, 但其频数仅为 5, 由此可见 ambiguous 一词多表现为中性语义韵。

从前 100 个显著搭配词中, 发现 obscure 的显著的中性搭配词有很多类, 例如法律条款类: law、provision、regulation 等; 位置类: location、village、towns、origin 等, 它还可以与表示原因、乡村、物体、话题等词语搭配。obscure 也有消极的语义韵, 它可以和 trivia、cult、diseases、corners、joke、dispute、backwater 搭配。不难发现 obscure 与消极词汇的搭配有 7 个, 与中性词汇的搭配有 91 个, 而与积极词汇搭配的语义韵只有 2 个, 因而 obscure 与 ambiguous 都没有明显的消极语义韵, 多表现为中性语义韵。由此可见, obscure 也是一个带有中性语义韵的词, 且与其搭配的词类别十分丰富。

5. 总结

语言技能、语言知识、情感态度、学习策略和文化意识等多维度的因素都能够影响同义词的学习(韩宝成, 2010 [6])。可能正是由于缺乏对这些因素的重视才会导致学生对近义词的错用与误用, 因此英语教学应该注重对词汇搭配和语义韵知识的传授。词的主要用法形式和典型组合是区别近义词搭配行为和语义韵特征的重要依据, 英语教师应考虑将这些非语言内容纳入外语教学。这将对激发学生对英语学习的兴趣, 培养学生良好的学习习惯, 加深学生对英语国家文化的了解具有重要意义。

近年来,国内基于语料库的外语研究逐步展开,我们应重视基于语料库工具的近义词教学。语料库已成为外语研究的新工具,为词典的编纂、翻译学和语言教学的应用研究提供了强有力的支撑。针对我国英语学习者因缺乏搭配行为辨别和语义韵知识而误用近义词的现象,有必要引导学生在日常学习中学会运用语料库工具,从而区分近义词搭配与语义韵特征差异。

参考文献

- [1] 张继东,刘萍.基于语料库同义词辨析的一般方法[J].解放军外国语学院学报,2005(6):49-52.
- [2] 杨慧中.语料库语言学导论[M].上海:上海外语出版社,2002:23-27.
- [3] 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.
- [4] Louw, B. (2000) Contextual Prosodic Theory: Bringing Semantic Prosodies to Life. In: Heifer, C., Sauntson, H. and Fox, G., Eds., *Words in Context: A Tribute to John Sinclair on His Retirement*, University of Birmingham, Birmingham, 48-67.
- [5] Stubbs, M. (1995) Collocations and Semantic Profiles: On the Cause of the Trouble with Quantitative Studies. *Functions of Language*, 2, 23-55. <https://doi.org/10.1075/fo1.2.1.03stu>
- [6] 韩宝成.关于我国中小学英语教育的思考[J].外语教学与研究,2010(4):300-302.