

基于COCA语料库的动词“Derive”的扩展意义单位研究

黄琴超

浙江工商大学外国语学院, 浙江 杭州

收稿日期: 2026年8月17日; 录用日期: 2026年9月9日; 发布日期: 2026年9月22日

摘 要

本研究基于美国当代英语语料库(COCA), 运用John Sinclair提出的“扩展意义单位”(EUM)模型, 对动词“derive”的用法进行了系统性分析。文章首先通过检索高频搭配词, 将其归纳为“收益类”、“知识与信息类”和“科技类”三个语义类别。随后, 文章识别出与“derive”相关的四种主要类联接。在此基础上, 研究进一步归纳出六种语义倾向, 并最终分析了“derive”的语义韵。研究发现, 动词“derive”在绝大多数语境中呈现积极的语义韵, 即与获益、满足、启发等正面概念相关联; 仅在如涉及不道德来源的收益或伦理争议的科学主题等少数特定语境下呈现消极的语义韵。文章旨在揭示词汇意义在真实语境中的构建过程, 并为二语习得提供启示。

关键词

扩展意义单位, 语料库, 语义韵, “Derive”

A Corpus-Based Study of the Extended Unit of Meaning of the Verb “Derive” in COCA

Qinchao Huang

School of Foreign Languages, Zhejiang Gongshang University, Hangzhou Zhejiang

Received: August 17, 2026; accepted: September 9, 2026; published: September 22, 2026

Abstract

Based on the Corpus of Contemporary American English (COCA), this study systematically analyzes the use of the verb *derive* within the framework of John Sinclair’s Extended Units of Meaning (EUM) model. First, the study identifies the high-frequency collocates of *derive* and categorizes them into three semantic groups: benefit-related, knowledge- and information-related, and science- and

文章引用: 黄琴超. 基于 COCA 语料库的动词“Derive”的扩展意义单位研究[J]. 现代语言学, 2026, 14(9): 591-597.

DOI: 10.12677/ml.2026.149959

technology-related terms. It then identifies four major types of colligations associated with *derive*. On this basis, six semantic preferences are further identified, followed by an analysis of the semantic prosody of *derive*. The findings show that the verb *derive* exhibits a predominantly positive semantic prosody in the majority of contexts, being associated with positive concepts such as benefits, satisfaction, and inspiration. A negative semantic prosody is observed only in a limited number of specific contexts, such as those involving benefits obtained from unethical sources or scientific topics involving ethical controversies. This study aims to reveal how lexical meaning is constructed in authentic contexts and to provide insights into second language acquisition.

Keywords

EUM, Corpus, Semantic Prosody, “Derive”

Copyright © 2026 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

随着语料库语言学的发展,基于真实语料进行语言的意义研究已成为语言研究的重要趋势。John Sinclair 提出的扩展意义单位模型为意义分析带来了独特的视角,它强调,语言的意义不仅存在于词典中,更通过人们实际交际使用中体现出来[1]。因此,采用扩展意义单位模型研究真实语境中的词汇使用规律,对于深入理解语言的意义具有重要价值。

近年来,扩展意义单位模型已被广泛应用于词汇语义和语用研究。例如,卫乃兴(2012)以汉语“展现”为节点词,运用该模型展示了语料库驱动的短语单位研究的主要特征和方法[2];Hunston (2011)则从理论层面深化了语义韵与评价语言之间的关系[3]。在针对具体词汇的研究中,刘运锋(2023)利用 COCA 语料库的具体实例讨论扩展意义单位再扩展分析的理论与其在词典编纂领域的应用价值[4]。这些研究均表明,EUM 模型对于揭示词汇在实际使用中的意义构建机制具有一定的解释力。

动词 *derive* 在英语中经常出现在各正式体裁中,尤其在科技、经济和哲学等领域的文本中频繁被使用。作为一个抽象动词,*derive* 具有较强的搭配依赖性,其实际使用中的意义构建更需要依托语料进行深入分析。基于此,本文拟采用 Sinclair 的扩展意义单位理论,在美国当代英语语料库(COCA)中检索,对动词 *derive* 在真实语境中的搭配词、类联接、语义倾向,以及语义韵进行系统分析其在语境中的具体应用,以期对二语学习者提供一定借鉴和参考。

2. 扩展意义单位概述

John Sinclair 于 1996 年首次提出扩展意义单位(Extended Units of Meaning)的概念,简称 EUM,它由五个部分构成:意核(core)、搭配词(collocation)、类联接(colligation)、语义倾向(semantic preference)以及语义韵(semantic prosody) [1]。

意核指的是研究考察的词,是寻找扩展意义单位的切入点。意核不能预先确定,亦不等同于节点词,只有在找到了扩展意义单位后,意核才得以确定[5]。搭配是由两个或多个词构成、共现频率较高的语言单位。Sinclair 将搭配定义为在意义单位中与意核习惯性共现而形成的词汇模式,旨在描绘意核与其周边成分的组合关系[6]。类联接则是意核与其周边成分之间的语法关系,该概念最早由 J. R. Firth (1957)提出[7],Sinclair 在其基础上进一步定义为“一个语法类或结构型式与其他语法类或结构型式,或其与某词或

短语的共现”[6]。语义倾向是意核周边的语义集合。如果意义单位的语义倾向一致，则可视作相同的扩展意义单位；若不一致，则需将这些扩展意义单位划分为不同类型。语义韵是整个意义单位传达的一般性的态度意义，在整个扩展意义单位中占统领地位[8]。

通过扩展意义单位的分析，研究者可以深入揭示词汇在真实语境中形成的意义网络，有效弥补词典释义的不足，进而更好地理解语言使用的常规性与创造性。

3. 研究方法

3.1. 研究问题

- (1) 在 COCA 语料库中，动词“derive”搭配与类连接方面有什么特点？
- (2) 在 COCA 语料库中，动词“derive”在不同的搭配和类连接中有哪些语义倾向以及语义韵？

3.2. 语料来源

本研究借助的语料库工具软件为 COCA 语料库。The Corpus of Contemporary American English (COCA) 为美国当代英语语料库，由杨百翰大学的马克·戴维斯教授开发，是美国具有代表性且使用较为广泛的大型英语语料库。它涵盖超过 10 亿字的文本，每年仍在不断更新，包括口语、小说、流行杂志、报纸、学术文本等 8 种文本类型，为本研究提供较为丰富的语料。

3.3. 研究步骤

本研究选择及物动词“derive”为节点词，运用扩展意义单位对其展开分析。首先，在 COCA 语料库界面检索动词“derive”的所有屈折形式，将跨距(span)设为-4/+4，以期考察其周围的搭配词，并列出词频最高的前 20 位搭配词。其次，结合索引行观察“derive”与其周边搭配成分之间的结构关系，在此基础上根据“derive”周边的词类、句法位置等结构性成分的共现情况，对类联接作归纳。再次，基于上述得出的类联接对“derive”的搭配词进行语义分类。分类主要依据搭配词所指称的对象及其在语境中的语义功能，并结合相近类联接中的共现情况进行归纳，从而确定不同类联接下“derive”的主要语义倾向。最后，在具体语境中考察整个扩展意义单位体现的态度意义，进而分析“derive”的语义韵。

4. 动词 Derive 的扩展意义单位分析

4.1. 搭配

正如 Firth (1968) 的著名论断“欲知其词，先观其伴”(You shall know a word by the company it keeps!)[9]，搭配分析在寻找扩展意义单位中的重要性不言而喻。研究为了考察 derive 的所有屈折形式，在 COCA 语料库的搜索栏中输入其大写形式，将跨距设为-4/+4，按照搭配频率列出前 20 个搭配词，如表 1 所示：

Table 1. Top 20 collocates of derive in COCA corpus

表 1. 动词 derive 在 COCA 语料库的前 20 个搭配词

序号	搭配词	频率	序号	搭配词	频率
1	benefit	576	11	source	283
2	word	528	12	pleasure	267
3	power	483	13	income	266
4	data	464	14	meaning	197
5	name	459	15	score	195

续表

6	cell	348	16	revenue	194
7	value	335	17	experience	185
8	model	318	18	estimate	184
9	study	308	19	satisfaction	181
10	information	286	20	energy	180

可以发现,前 20 个搭配词均为名词,根据它们的语义特征可细分为 3 组。第一组是“收益类名词”: benefit, pleasure, satisfaction, income, revenue, experience。它们大多表示某种可以被获得的结果,包括物质收益、主观体验以及由某种行为所带来的结果等。例如, derive benefit from、derive income from 和 derive satisfaction from 均表现出较为清晰的“获得-来源”关系。第二组是“知识与信息类名词”: word, name, meaning, source, data, information, model, study, estimate, 这类搭配词主要涉及信息、知识、概念及其来源。其中, word, name, meaning 更多与词汇、术语及概念的来源有关,而 data, information, model, study 等则更多出现在学术或专业语境中。这也在一定程度上说明了动词 derive 在学术文本中的重要性。第三组是“科技类名词”: power, energy, cell, score, 多用于如物理、工程、生物等领域的科技类文本,强调“通过某种机制生成或获得某种实体或特性”。这类搭配词的具体意义则更加依赖语篇所处领域。

总的来看, derive 的高频搭配词虽然涉及收益、知识信息和科技等不同语义领域,但多数都可以与“获得、产生或追溯来源”这一较为概括的语义联系起来。但是,仅从单个词的词典释义出发,较难充分呈现 derive 在实际语境中的意义差异,因此,还需要结合其类联接及后续的语义倾向进行进一步考察。

4.2. 类联接

类联接指的是意核周边所构成的语法型式(grammatical patterns),是寻找扩展意义单位的第三步。当我们依次观察 derive 左一、右一位置上的搭配词发现,不定式 to (1232)、情态动词如 can (186), may (129)、否定词 not (166)以及一些副词如 directly (91), much (82), primarily (60), largely (56)等具有较高的复现频率,但是濮建忠(2021)不建议把此类情态动词、程度副词纳入扩展意义单位之中,它们存在于普通语法中 [10]。因此,本研究通过进一步观察意核与其高频搭配词的索引行,发现与动词 derive 有关的类联接主要有以下 4 种,如表 2 所示:

Table 2. Distributions of colligation of derive in COCA corpus

表 2. Derive 在 COCA 语料库中的类联接分布

类联接	索引行示例
DERIVE + NOUN	we can derive a new algorithm that satisfies this rule... The Federal Reserve is an independent central bank that derived its power in the aftermath of the Panic of 1997...
DERIVE + NOUN + FROM + NOUN	...the AMC is deriving excessive profit from its hut operationsthey're not going to derive some sort of income from working in the place...
BE + DERIVED + FROM + NOUN	...no excuse can be derived from the ignorance of where the sovereignty is placed. This model is derived from the analytic tradition in philosophy.
DERIVE + FROM + NOUN	...the flag may derive from a crusade banner or ensign note... Many of the most serious issues derive from data ...

仔细考察索引行发现，“derive”既可以作及物动词，也可以作不及物动词，前3个类联接都属于及物动词的用法。当它后面仅接一个宾语时，通常用于表达动作发出者通过某种过程“获得”某种结果；在该结构中，宾语直接表示“获得”的结果，而其具体来源往往需要结合上下文才能确定。而当它后面再接介词 from 时，则进一步明确获得的结果的来源，使整个句式结构和逻辑都更完整。

与此同时，“derive”也经常用于被动语态中，它突出了来源的重要性，并弱化了动作的发出者，常见于学术与科技文本中。与主动结构相比，被动结构弱化了动作执行者，突出结果与来源之间的关系，因此在需要客观陈述来源、依据或形成过程的语境中具有较强的适用性。此外，derive 作不及物动词时接介词 from，这里的 derive 不同于前面三种，表达主语的起源或归属，常用来表示语言、文化、术语等背景来源，故介词 from 不可省略。

4.3. 语义倾向

语义倾向是 Sinclair 提出的概念，是描写与意核搭配的一些实词(或词组)的语义特征，在同一个扩展意义单位中，一些与意核搭配的实词尽管词类不尽相同，但在语义上可视为同一类别[5]。基于前面的搭配以及类联接分析，结合 derive 的索引行，可以归纳其以下几种语义倾向：

Table 3. Classifications of semantic preference of derive in COCA corpus
表 3. Derive 在 COCA 语料库中的语义倾向分类

语义倾向	搭配词
intellectual acquisition	data, information, knowledge, insights, hypotheses, conclusions, estimates, models, measures
material benefit	income, revenue, revenues, benefit, benefits, profits, wealth, advantage, value, values
linguistic and etymological origin	name, word, term, title, meaning, verb, Latin, Greek, Arabic
related to scientific, logical, biological and natural issues	empirically, logically, principles, factors, variables, equations, formula, models, stem cells, vaccines, drugs, compounds, plants, embryonic, herd, immunity, energy
subjective experience	pleasure, satisfaction, inspiration, hallucinations, images, meaning, strength, legitimacy

表 3 清晰地告诉我们，动词 derive 在 COCA 语料库中的语义倾向主要有 6 类。第一类主要是和知识、信息相关的词，常出现在“DERIVE + NOUN”和“DERIVE + NOUN + FROM + NOUN”中，用来表达通过分析或推理获取一些抽象的认知产物，例如“**derive information** about gravity”。第二类表达的意思是个体或者群体获取物质上的收益，通常涉及利益、收入、价值等，例如“**derive income from** the work”，“**Governments derive their just powers from** the consent of the governed”。第三类是和词汇、术语或概念有关的词，用来表达词汇的来源，因此它高度依赖介词 from，如“**a name derived from** the word ubiquitous”。第四类搭配词与数学逻辑、自然科学息息相关，常见于“BE + DERIVED + FROM + NOUN”，通过被动语态弱化行为主体，突出自然本源属性，比如“**compounds derived from** ribose and deoxyribose”。第五类是一些主观体验的搭配词，用来表达人们在情感、心理或认知层面的非物质层面的获得，如“**all adults derive satisfaction from** a basic knowledge of science”。

4.4. 语义韵

不同于语义倾向(实义搭配词的语义特征)，语义韵是整个意义单位所传达的态度意义，由意核与各类搭配词所组成的复合词项整体所决定，是该复合词项蕴含的、不言而喻的一般性态度意义[5]。基于上述的类联接与语义倾向的分析，我们可以总结出 derive 的语义韵。在 COCA 语料库中，derive 主要呈现出

积极的语义韵，偶见消极的语义韵。

4.4.1. 积极的语义韵

结合各类联接与语义倾向观察索引行发现，当 *derive* 与在“物质层面的收益”、“主观体验”这些语义倾向搭配时，其语义韵呈现明显的积极性。在“DERIVE + NOUN + FROM + NOUN”类联接中，如果动词接的宾语是 *benefits* (收益)、*value* (价值)等指向集体福祉的名词，且其来源具有社会正当性，说明动作发出者通过合理的渠道获益(见例 1、例 2)。人们通过某个过程获得满足感(*satisfaction*)、愉悦(*pleasure*)、启发(*inspiration*)等都表示出正向的态度意义，常见于描述理性分析、学习过程或制度安排等带来的收获与正向体验，有时其周边还有程度副词进一步地强化该积极韵(见例 3、例 4)。在一些学术文本中，*derive* 和“知识获取”类实义搭配词共现时，也形成了隐性的积极的语义韵。通过“BE + DERIVED + FROM + NOUN”这一被动语态的结构为结果提供了可信度，这种积极的评价不依赖情感词汇，而是源于自然科学话语“可推导性”的价值共识(见例 5)。

例 1: ...private landowners do **derive benefit from some of the services** provided by the ecosystems on their land...

例 2: Embedding Pentaho into Creditability is helping us to **derive much more value from our main business assets**...

例 3: ...because they see a need for the work to be done and they **derive satisfaction from meeting that need**.

例 4: These articulations **derive significant insight from** womens labour and everyday networks at the local level, ...

例 5: Most of our **knowledge is derived from** experimental animal models.

4.4.2. 消极的语义韵

然而，值得注意的是，*derive* 在少数的语境中显现出了消极的语义韵。最典型的是，我们发现当 *derive* 后接宾语 *hallucinations* (幻想)、*paranoia* (偏执)等病态心理时，因这些词不再表示正向的获得，而是某种负面的产物。在“物质层面的收益”这个语义倾向中，如果“DERIVE + NOUN + FROM + NOUN”的宾语是指向私利的词，如 *profits* (利润)、*gains* (利益)等，且介词 *from* 后接“垄断”(monopolies)等有伦理争议的行为，那消极的语义韵便会显现出来(见例 7)。同样的消极韵也可能出现在“自然科学问题”语义倾向中，如“BE + DERIVED + FROM + NOUN”中的主语涉及“胚胎”(embryos)、“动物组织”(animal tissues)等有关生命伦理的对象时，语义韵明显转向负面(见例 6)。

例 6: Scientists have devised two new techniques to **derive embryonic stem cells in mice**...

5. 结语

本文借助 COCA 语料库分析了动词 *derive* 的扩展意义单位，分别对其搭配、类联接、语义倾向和语义韵的特点进行了系统的分析。结果表明，动词 *derive* 主要与其搭配词构成了 4 种类联接，并在此基础上主要有 5 种不同的语义倾向，最后，*derive* 绝大多数情况下体现出积极的语义韵，也有少量的消极的语义韵。究其本质，扩展意义单位是一个整体，是各个成分之间的共选关系。通过探寻扩展意义单位，我们可以从词汇搭配、语法结构及语用的角度来考察单词，单词的意义便不能仅是通过单词本身来识别。可以说，扩展意义单位的研究为二语习得提供了新的视角与路径，教学与习得中都应该结合语境掌握单词的搭配与用法，避免孤立地记忆词义，进而提高语用能力。

参考文献

- [1] Sinclair, J.M. (1996) The Search for Units of Meaning. *Textus*, 4, 75-106.
- [2] 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, 125.
- [3] Hunston, S. (2011) *Corpus Approaches to Evaluation: Phraseology and Evaluative Language*. Routledge.

<https://doi.org/10.4324/9780203841686>

- [4] 刘运锋. 扩展意义单位再审视[J]. 外国语(上海外国语大学学报), 2023, 46(4): 17-26.
- [5] 濮建忠. 扩展意义单位模型再解读[J]. 外语研究, 2020, 37(2): 1-8, 112.
- [6] Sinclair, J.M. (2003) *Reading Concordances: An Introduction*. Pearson.
- [7] Firth, J.R. (1957) *Papers in Linguistics: 1934-1951*. Oxford University Press.
- [8] 李文中. 接着做: 扩展意义单位分析[J]. 当代外语研究, 2021(6): 13-26, 88, 161.
- [9] Firth, J.R. (1968) *Descriptive Linguistics and the Study of English*. In: Palmer, F., Ed., *Selected Papers of J. R. Firth 1957-1959*. Indiana University Press, Bloomington, 84-95.
- [10] 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, 160.