

# COCA Corpus-Based Study of the Near-Synonymous Verbs

## —Taking “Destroy”, “Damage” and “Ruin” for Reference

Shumin Zhuo

Wuhan University of Technology, Wuhan Hubei  
Email: 490536058@qq.com

Received: Jan. 24<sup>th</sup>, 2019; accepted: Feb. 7<sup>th</sup>, 2019; published: Feb. 14<sup>th</sup>, 2019

---

### Abstract

With the rise of corpus linguistics, natural language data in the corpus have gradually become an important tool in language teaching. Based on COCA corpus, this paper tries to discuss the differences among the three near-synonymous verbs “destroy”, “damage” and “ruin”, and makes a comparative analysis of them in word frequency, stylistics, collocation and semantic prosody. The purpose of this study is to discriminate these near-synonymous verbs and promote English learners' interest in using corpus data.

### Keywords

COCA Corpus, Near-Synonymous Verbs, Frequency, Stylistics, Collocation, Semantic Prosody

---

# 基于COCA语料库的近义动词辨析

## ——以Destroy, Damage和Ruin为例

卓淑敏

武汉理工大学, 湖北 武汉  
Email: 490536058@qq.com

收稿日期: 2019年1月24日; 录用日期: 2019年2月7日; 发布日期: 2019年2月14日

---

### 摘要

随着语料库语言学的兴起, 语料库中的自然语言数据逐渐成为语言教学中的重要工具。本文将基于COCA

语料库, 探讨“damage”, “destroy”和“ruin”这三个近义词的异同, 对其在词频、文体、搭配和语义韵这四方面进行对比分析。目的是更加准确地区分这组近义词, 同时引起英语学习者对语料库的使用兴趣。

## 关键词

COCA语料库, 近义词, 词频, 文体, 搭配, 语义韵

Copyright © 2019 by author(s) and Hans Publishers Inc.

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

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



Open Access

## 1. 引言

由于语言的任意性和创造性, 近义词在英语中占有很大比例。据统计, 英语中同义词、近义词的数量约占总词汇量的 60% [1]。对近义词的辨析一直是外语学习中一个难以掌握而又至关重要的环节。近义词虽然基本表达相同的概念, 但是在不同的文体、上下文里、从不同角度看来还是有所差别的。在传统教学中, 近义词辨析通常有如下几种方法: 教材注解、词典释义和惯用搭配。这些方法较为死板且缺乏真实的语境支撑, 无法解决深层的语义区分问题, 学生们也难以把握近义词准确地道的用法。

语料库语言学的逐步发展为近义词辨析提供了新的思路。语料库是指按照一定的语言学原则, 运用随机抽样方法, 收集自然出现的连续的语言文本或话语片段而建成的具有一定容量的大型电子文库[2], 具有容量大、语料真实、检索快捷的独特优势[3]。运用语料库对相关词汇进行检索、对比和分析, 可以对近义词进行根本的语义区分。

国内外的学者在基于语料库的近义词辨析方面取得了较丰硕的成果。1996年, Hanks 第一个使用词项整体行为研究(behavioral profile)的方法研究了“urge”, “incite”, “bother”和“abandon”这四个近义词, 他所采用的是 BNC 语料库(British National Corpus) [4]。随后, Dilin Liu 采用同样的方法, 基于 COCA 语料库研究“chief”, “main”, “major”, “primary”和“principal”所修饰的名词类型, 探究各个形容词的内部语义结构[5]。郭放、翟红华对比研究了“affect”的类联接模式[6]。梅一馨探究了含有“影响”含义单词的词性转换对其语义韵的影响[7]。但基于语料库方法辨析近义词的研究大都限于研究各近义词的类联接、搭配选择和语义韵中的某个方面, 较少同时考察近义词在上述各方面的差异。

本文采用现代美国英语语料库(the Corpus of Contemporary American English, 简称 COCA), 旨在研究“destroy”, “damage”和“ruin”这三个近义词在语义和用法上的不同。COCA 语料库由杨伯翰大学的 Mark Davies 教授开发, 是当今世界最大的美国英语平衡语料库。截止到 2017 年 12 月, COCA 语料库共收集 5.6 亿词, 涵盖 1990 年至 2017 年期间美国境内口语、小说、杂志、报纸及学术期刊五大领域的语料。基于 COCA 语料库, 本文同时从词频、文体、搭配和语义韵这四个角度来研究“destroy”, “damage”和“ruin”这三个近义词的不同, 帮助学习者系统地地区分近义词, 为语言教学中的近义词辨析提供参考。

## 2. 基于 COCA 语料库辨析近义词“destroy”, “damage”和“ruin”

《朗文当代英语词典》(Longman Dictionary of Contemporary English)对“destroy”, “damage”和“ruin”这三个词的解释如下:

destroy: to damage sth. so badly that it no longer exists, works, etc.

damage: to cause physical harm to sth., making it less attractive, useful or valuable; to have a bad effect on sb./sth.'s life, health, happiness or chances of success.

ruin: to damage sth. so badly that it loses all its value, pleasure, etc. [8].

根据以上释义, 若仅仅学习独立的目标词, 学习者可借由这些信息对这三个词进行良好掌握; 然而一旦对该组近义词进行对比, 区分难度便随之而来——如上信息向我们展示的词义大有重叠之处, 甚至是用一个单词来解释另一个单词。这样非常不利于英语学习者的学习。因此, 基于语料库的近义词辨析还是很有必要的。下文将从词频、搭配、文体和语义韵这四个角度对这三个近义动词进行分析。

## 2.1. 词频

本文通过 COCA 语料库分别检索了“destroy”, “damage”和“ruin”这三个近义动词, 选择图表分析后得到它们的总频数。由于这三个词在英语中都具有一种以上的词性, 因此本文在 COCA 的搜索引擎中每个动词的后面加上了“.[v\*]”。

**Table 1.** Word frequency of the three verbs

**表 1.** 三个动词的词频

单词	词频
destroy .[v*]	103,288
damage .[v*]	36,576
ruin .[v*]	27,704

从表 1 的数据中, 我们可以看到出现频率最高的是“destroy”, 出现频率最低的是“ruin”。按照表 1 中的数据, 我们可以把这三个动词从高到低排序: destroy→damage→ruin。也就是说, “destroy”在英语中最常使用, 用来表达损害或毁坏某件物体, 而“ruin”使用频率最低。虽然总体频率并没有告诉我们任何关于每个动词的意思和特殊的使用模式, 但是对于英语学习者来说, 当他们想表达损害或毁坏某件物体时, 他们可以更多地使用“destroy”和“damage”这两个动词。

## 2.2. 文体

词汇的分布模式对于语言学习者来说非常重要。不同文体内, 近义词分布频率不同, 统计不同文体的词频差异能够促进学习者文体意识的培养, 发现在特定文体用词优先原则, 能宏观上指导近义词辨析。因此这三个近义动词的文体研究对于英语学习者来说相当有意义。本文通过 COCA 语料库的图表显示检索了“destroy”, “damage”和“ruin”在不同文体中的出现频数和每百万词的频率, 得出了表 2 的数据。

**Table 2.** Word frequency in different styles

**表 2.** 不同文体中的词频

.[v*]	destroy		damage		ruin	
	词频	频率(permil)	词频	频率(permil)	词频	频率(permil)
spoken	25,344	217.08	6392	54.75	5384	46.12
fiction	21,096	188.62	3856	34.48	10,144	90.70
magazines	23,240	198.03	12,528	106.75	6000	51.13
newspapers	17,552	155.33	7328	64.85	4504	39.86
academic	16,056	144.12	6 472	58.09	1672	15.01

上述数据表明：1) 在各个文体中，“destroy”的使用均高于“damage”和“ruin”，而且在各个语域的词频波动不大，表明“destroy”在各个语境都适用；2) “damage”在报纸这一文体中词频高达 12,528，远远超过在其他四个语域中的词频，这表明该词适合用于官方报道；3) “ruin”一词在学术文体中使用频率最低，表明“ruin”一词常用于非正式文体。由此可见，通过调查近义词在不同文体中的分布频率，可以发现近义词在不同文体中的优先用词原则，以便用语更加地道。

### 2.3. 搭配

搭配，即词的典型组合，是指在文本中实现一定的非成语意义并以一定语法形式因循组合使用的一个词语序列[9]。Firth 曾指出“You shall know a word by the company it keeps” [10]。也就是说，通过搭配词，我们可以很好地对一个词进行理解和研究。本文通过 COCA 对近义词“destroy”，“damage”和“ruin”进行搭配检索，寻找这三个词前后出现频率较高的词。同一个搭配词出现的频率越高，表明其与目标词的关联性越强，搭配强度越大。本文还将利用 MI 值为参考对搭配强度进行分析。所谓 MI 值(Mutual Information Score)就是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，那就意味着这两个词基本不搭配在一起；如果 MI 值大于 3 或更高，就意味着这两个词经常在一起出现，也就是互相搭配。MI 值越大，说明节点词对其词汇环境影响越大，对其共现词吸引力越强。

#### 2.3.1. 与介词搭配

众所周知，动词常常和介词搭配。在之前提到的词典当中，这三个动词均不和介词搭配使用。在语料库中这三个单词是不是也是如此使用？作者在“COLLOCATION”一栏中输入“[i\*]”，把左边搭配的数量降低为 0，右边的搭配数量设定为 4，所有的数据都按“RELEVANCE”从大到小排列，这样就能得出这三个近义词与各个介词搭配的频率和 MI 值。表 3 因为篇章有限只列出了前 5 个 MI 值最高的与这三个近义词搭配的介词。

由表 3 可见，三个近义词与各个介词的相关性都不高，大部分搭配的 MI 值在 0 左右，甚至低于 0，也就是说这三个近义词几乎不与介词搭配。参见 COCA 语料库所给例句：

- 1) We will also destroy those defense systems without hesitation.
- 2) What if you damage it at all during the process?
- 3) I wasn't going to ruin that progress by greeting new clients without his permission.

从以上例句可以看出，虽然这三个动词会与一些介词同时出现在同一个句子中，但它们并不和介词搭配。

Table 3. The first five prepositions collocated with the three verbs

表 3. 前 5 个分别和三个动词搭配的介词

	destroy .[v*]		damage .[v*]		ruin .[v*]				
1	WITHOUT	Freq	44	DURING	Freq	20	BY	Freq	87
		MI	1.00		MI	1.17		MI	0.76
2	WITHIN	Freq	23	BY	Freq	101	FOR	Freq	180
		MI	0.62		MI	0.58		MI	0.60
3	BY	Freq	204	FROM	Freq	109	WITH	Freq	64
		MI	0.09		MI	0.54		MI	-0.55
4	OF	Freq	1259	TO	Freq	540	LIKE	Freq	19
		MI	-0.08		MI	0.21		MI	-0.57
5	ALONG	Freq	12	BETWEEN	Freq	11	OF	Freq	223
		MI	-0.31		MI	-0.15		MI	-0.68

### 2.3.2. 与名词搭配

在词典中，“destroy”是及物动词，“damage”和“ruin”既可以作为及物动词也可以作为不及物动词。为了降低可变性、更科学地比较这三个近义词，笔者将这三个动词在都作为及物动词的情况下进行比较。将“[nm\*]”输入“COLLOCATION”一栏中，左边的单词数量设为0，右边的单词数量设为4，所有的数据都按“RELEVANCE”从大到小排列。并且所有的搭配名词都以“LEMMAS”的形式表现出来，即把一个名词的单数和复数形式归为一个词条。

为了更好地分析数据，笔者将这三个近义词所搭配的名词分为具体名词和抽象名词。从这些数据中我们可以得出，“destroy”多和具体名词搭配，且与战争有关的名词多和“destroy”搭配，例如，“warhead”、“missile”、“enemy”、“arsenal”等。和“damage”搭配的名词大部分为具体名词，且在这些具体名词中，与身体构造相关的名词频频出现，如“kidney”、“artery”、“liver”、“nerve”、“organ”、“DNA”、“vessel”、“lung”、“brain”等。“ruin”多和抽象名词搭配，很少和具体名词搭配。例如，“reputation”、“appetite”、“friendship”、“happiness”等。下面的三个图表(表4~6)因为篇章有限只列出了前20个MI值最高的与这三个近义词搭配的名词。

**Table 4.** The first 20 nouns combined with “destroy”

**表 4.** 前 20 个和 “destroy” 搭配的名词

	Nouns	Freq	MI		Nouns	Freq	MI
1	[OZONE]	55	7.01	11	[ENEMY]	109	5.26
2	[EMBRYO]	35	6.55	12	[HABITAT]	48	5.26
3	[WARHEAD]	14	6.23	13	[CIVILIZATION]	38	5.20
4	[LIVELIHOOD]	16	6.13	14	[REPUTATION]	59	5.12
5	[MICROORGANISM]	10	6.09	15	[ARSENAL]	12	5.05
6	[STOCKPILE]	12	5.91	16	[MORALE]	15	5.01
7	[CREDIBILITY]	50	5.90	17	[BACTERIA]	34	4.95
8	[WETLAND]	24	5.70	18	[ENZYME]	10	4.76
9	[MISSILE]	79	5.47	19	[TUMOR]	25	4.75
10	[MANKIND]	12	5.31	20	[INFRASTRUCTURE]	32	4.74

**Table 5.** The first 20 nouns combined with “damage”

**表 5.** 前 20 个和 “damage” 搭配的名词

	Nouns	Freq	MI		Nouns	Freq	MI
1	[AVERAGING]	10	9.33	11	[NERVE]	38	6.37
2	[LINING]	13	7.45	12	[ORGAN]	22	5.84
3	[REPUTATION]	88	7.20	13	[DNA]	33	5.80
4	[KIDNEY]	27	6.97	14	[VESSEL]	25	5.66
5	[CREDIBILITY]	35	6.88	15	[SELF-ESTEEM]	12	5.57
6	[OZONE]	16	6.73	16	[CROP]	32	5.55
7	[ARTERY]	16	6.64	17	[LUNG]	21	5.55
8	[LIVER]	24	6.54	18	[INTEGRITY]	15	5.32
9	[MEMBRANE]	12	6.52	19	[ECOSYSTEM]	13	5.20
10	[TISSUE]	50	6.45	20	[BRAIN]	59	5.14

**Table 6.** The first 20 nouns combined with “ruin”  
**表 6.** 前 20 个和 “ruin” 搭配的名词

	Nouns	Freq	MI		Nouns	Freq	MI
1	[REPUTATION]	81	7.48	11	[WEDDING]	12	4.34
2	[APPETITE]	19	6.99	12	[MARRIAGE]	26	4.30
3	[CAREER]	111	5.90	13	[CHRISTMAS]	14	3.90
4	[FRIENDSHIP]	17	5.66	14	[EVENING]	17	3.62
5	[VACATION]	20	5.53	15	[CLOTHES]	10	3.49
6	[HAPPINESS]	11	5.48	16	[ECONOMY]	23	3.48
7	[MOOD]	17	5.18	17	[FUN]	14	3.40
8	[LIFE]	318	4.81	18	[TRIP]	16	3.38
9	[CHANCE]	52	4.58	19	[RELATIONSHIP]	28	3.26
10	[HOLIDAY]	14	4.37	20	[NEIGHBORHOOD]	11	3.18

## 2.4. 语义韵

Louw 在 1993 年正式运用“语义韵”这一术语，他把语义韵定义为“一个词汇的意义的氛围受它的搭配影响”[11]。简单来说，语义韵即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。词语搭配形成的语义韵大体上可分为消极语义韵、中性语义韵和积极语义韵三类[12]。近义词辨析中，语义韵概念的引入对于确定近义词的态度标识功能起到了很好的解释作用[13]。

由 COCA 语料库所提供的数据可见(以表四、表五和表六为例)，三个近义词的搭配名词基本没有表现出积极的语义韵。“destroy”的搭配名词有一部分表达出中立的语义韵，如“microorganism”、“bacteria”、“enzyme”等，另一部分表达出消极的语义韵，如“livelihood”、“civilization”等。而与“damage”、“ruin”的搭配名词绝大部分表达出消极的语义韵，比如，它们均可与“reputation”等搭配。

## 3. 总结

文章通过语料库研究方法，借助 COCA 语料库，对近义词“destroy”，“damage”和“ruin”进行了出现频数、文体分析、搭配词和语义韵等四个方面的研究。得出如下结论：首先，“destroy”的词频在总体上最高，这意味着“destroy”是我们在表达损害或毁坏某件物体时最常用的一个单词。其次，从文体的角度来看，“destroy”在各个语境都适用，“damage”多用于官方报道，而“ruin”常用于非正式文体。第三，“destroy”、“damage”和“ruin”这三个近义词都几乎不和介词搭配。“destroy”多和具体名词搭配，且与战争有关的名词多和“destroy”搭配，和“damage”搭配的名词大部分为具体名词，其中绝大多数和身体结构有关，“ruin”多和抽象名词搭配。最后，这三个动词基本上没有表现出积极的语义韵。总而言之，语料库研究方法为近义词的区分提供了高效、快捷、准确的方式。对于英语学习者在英语择词上的具体考量起到了极大的帮助作用，而这一帮助也是极具说服力的。

## 参考文献

- [1] 贺晓东. 英语同义词词典[M]. 北京: 商务印书馆, 2003.
- [2] 杨慧中. 语料库语言学导论[M]. 上海: 上海外语教育出版社, 2002.
- [3] 杨春霞. 基于语料库的近义词辨析研究——以 suspect 和 doubt 为例[J]. 西南科技大学学报: 哲学社会科学版, 2014, 10(5): 45.
- [4] Hanks, P. (1996) Contextual Dependency and Lexical Sets. *International Journal of Corpus Linguistics*, 1, 75-98. <https://doi.org/10.1075/ijcl.1.1.06han>

- [5] Dilin, L. (2010) Is It a Chief, Main, Major, Primary, or Principal Concern? A Corpus-Based Behavioral Profile Study of the Near Synonyms. *International Journal of Corpus Linguistics*, **15**, 56-87. <https://doi.org/10.1075/ijcl.15.1.03liu>
- [6] 郭放, 翟红华. 基于语料库的“affect”的类联接对比研究[J]. 语文学刊(外语教育教学), 2012(10): 19-20.
- [7] 梅一馨. 基于语料库关于词性转换对语义韵影响的研究——以 affect/effect 和 influence 为例[J]. 兰州教育学院学报, 2015(1): 97-99.
- [8] 英国朗文出版公司. 朗文当代英语词典[M]. 北京: 世界图书出版公司, 1993.
- [9] 卫乃兴. 专业性搭配初探——语料库语言学的方法[J]. 解放军外国语学院学报, 2001, 24(4): 19-23.
- [10] Firth, J. (1957) Modes of Meaning. In: Firth, J., Ed., *Papers in Linguistics 1931-1951*, Oxford University Press, New York.
- [11] Louw, B. (1993) Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies. In: Francis, M. and Tognini-Bonelli, E., Eds., *Text and Technology: In Honor of John Sinclair*, John Benjamins, Philadelphia and Amsterdam. <https://doi.org/10.1075/z.64.1.1lou>
- [12] 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>
- [13] 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, 28(6): 53-56.

#### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2330-1708, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [ml@hanspub.org](mailto:ml@hanspub.org)