

基于语料库的近义词辨析研究——以 almost 和 nearly 为例

易倩文

武汉生物工程学院

DOI:10.12238/er.v8i5.6022

摘 要：近义词辨析一直是中国学生英语学习道路中的难点，而传统的近义词辨析教学一般是通过词典释义来进行的，但这一方法从实际而言无法让英语学习者对近义词的细微差异产生全面的认识。本文以副词 almost 和 nearly 为例，本文借助英语国家语料库(BNC) 从词频分布、搭配类连接和语义韵四个方面对 almost 和 nearly 的用法进行比较，旨在帮助英语学习者提供区分这两个近义词的方法，揭示这一对近义词的搭配和语义韵等特征，为近义词辨析提供新的视角。

关键词：近义词；搭配；类联接；语义韵

A Corpus-based Study of Near-synonym Identification
---Taking Almost and Nearly as an Example

Qianwen Yi

Wuhan University of Bioengineering

Abstract: Near-synonym identification has always been a difficult part of Chinese students' English learning path, and traditional teaching of near-synonym identification is usually carried out through dictionary explanations, but this method is not able to provide English learners with a comprehensive understanding of the nuances of near-synonyms in practical terms. In this paper, we take the adverbs almost and nearly as an example. With the help of the British National Corpus (BNC), this paper compares the usage of almost and nearly in terms of word frequency distribution, collocational connections and semantic rhymes, with the aim of helping learners of English to provide a way to differentiate between these two near-synonyms, revealing the collocational and semantic rhymes of these two pairs of near-synonyms, and providing new perspectives on near-synonym identification. The purpose of the comparison is to help English learners provide a way to distinguish between these two near-synonyms and reveal the features of collocation and semantic rhyme, providing a new perspective for near-synonym analysis.

Keywords: near-synonyms, collocation, class association, semantic rhyme

一、引言

英语是世界上使用最广泛的语言之一，词汇量十分丰富。语言学习中的近义词的辨析是一个难点。它通过词义、用法等方面的相同或相近构成了庞大而又复杂的英语词汇体系。陆军(2010)指出英语近义词在搭配行为和语义韵特征上存在差异,学习者在区别这些特征方面存在困难^[1]。就实际而言，近义词之间的细微差异给英语学习带来了挑战，我国传统的教学方式一般通过查阅字典来辨析近义词，但其只能提供词语解释，无法提供充足的词汇使用信息来辨别细微差异。Leech 曾经提出意义理论及成分分析法、框架语义学、意象图示等理论来深入分析近义词，但由于这些理论自身适用范围有局限性，只适用于部分近义词的辨析而不能囊括大部分

词汇。吴小芳（2016）基于认知语言学的范畴化理论和识解理论探讨了英语动词近义词的辨析，提出了英语动词近义词辨析的认知机制，为近义词辨析打开了新的思路。经查阅 almost 和 nearly 在《当代英语大辞典》(2004)中作为动词的解释，并进行了比较（见表 1）。

表 1 《当代英语大辞典》(2004)almost 和 nearly 释义

释义	例子
Almost 几乎，差不多，将近	1.I almost drop the plate. 2.She said almost nothing. 3.It's almost certain to succeed.

Nearly 1.几乎， 差不多， 将近	1.The train was nearly full.
2.密切地	2.A question which concerns me nearly.

从表一我们可以看出，尽管 nearly 的词义范围更大，但就其基本词义“几乎，差不多，将近”而言，两者的差别甚微。即使词典给出了具体语境，实际应用中也较难区分因此，就英语学习者而言，真实语境的例子也很难使学习者真正掌握近义词微妙的差异和合适的用法。此文采用语料库的研究方法，利用语料库的大规模的真实语料进行相关检索，并对检索结果加以分析总结，得出 almost 和 nearly 作为一组近义词动词的细微区别，给教师教学提供一点参考。

二、BNC 语料库

英语国家语料库（British National Corpus，简称 BNC）是由英国牛津出版社、朗文出版公司等联合开发建立的大型语料库。英国国家语料库（BNC）收录了 1 亿字的电子资源，用以呈现 20 世纪后期以来的英式英语，其涉及的范围较为广泛，其中书面语语料库达 90%，口语语料库达 10%。该语料库是当代英语规模大，代表性强的语料库。它全面、均衡地收集了各学科语域的语料，具有显著的代表性和可比性。该语料库对所有的词汇都作了标注，便于计算机进行检索分析和统计。在应用方面，BNC 语料库对使用者免费开放。BNC 语料库为研究者提供了庞大的原始语言资料，为客观的数据分析提供了可能。

三、研究设计

研究选取 almost 和 nearly 两个词主要基于以下两方面的考虑：第一，副词是中国学生学习的难点之一，词义的细微差别给英语学习带来了难题。第二，外语学习者缺乏有效的近义词辨别方法。英语本族语者大都能凭自己的直觉判断多义词的常用义项和搭配，但对于外语学习者而言，直觉判断往往不够准确。当几个多义词的义项十分相近时，外语学习者经常在使用中混淆它们。杨春霞（2014）认为基于语料库的近义词辨析借助语料库容量大、语料真实、检索快捷的特点，从词频、语域、搭配及语义韵特征等方面对近义词进行全方位辨析，为近义词辨析教学提供了新的视角^[3]。本文所选用的语料均来自于英语国家语料库（British National Corpus），作者对该语料库相关语料的检索，从而对 almost 和 nearly 在不同语域中的词频分布差异进行统计，揭示出这一对近义词的类连接和语义韵等语言特征。本文研究问题共有三个：第一，almost 和 nearly 在不同语域中的词频有什么差异？第二，almost 和 nearly 有哪些显著的搭配词？第三，almost 和 nearly 在 BNC 语料库中类连接及语义韵方面有何不同？

四、基于 BNC 的近义词 almost 和 nearly 分析

（一）almost 和 nearly 在 BNC 语料库不同语域中的分

布情况

语域（register）是人们在实际的语言活动中，出于交际的需要，或者因从事职业和兴趣的差异，话语的发生情景、说话对象、地点和话题的不同而产生的一种言语变体，通常体现为语言中的不同语体风格、用语格调等。近义词在不同的语域中会呈现出不同的分布特征。统计它们不同语域中的频率差异，分析分布特点，有助于英语学习者区分 almost 和 nearly 这一对近义词。首先，我们对 almost 和 nearly 在 BNC 中的整体分布频率来进行探究。

表 2 almost 和 nearly 在 BNC 中的词频分布

	总频数	每百万词频数
almost	30041	311.95
nearly	11005	114.28

通过观察以上数据可以发现：从总频数及每百万次频数来看，almost 较 nearly 而言使用的更频繁。尽管 almost 和 nearly 同样作为副词使用，但总体来看，almost 在 BNC 中的使用频率还是远远高于 nearly。接着，我们分别检索副词 almost 和 nearly 在不同语域中的词频分布情况，见表三和表四。

表 3 副词 almost 在不同语域中的词频分布

	口语	小说	杂志	报纸	非学术	学术	音乐
频数	1220	8212	2468	3305	4846	3916	6074
每百万词频	122.44	516.18	339.85	315.77	293.78	255.42	291.53

表 4 副词 nearly 在不同语域中的词频分布

	口语	小说	杂志	报纸	非学术	学术	音乐
频数	1312	2163	866	1684	1725	1096	2159
每百万词频	131.68	135.96	119.25	160.90	104.58	71.49	103.62

上述数据表明，总体而言，almost 的使用频率在各语域比 nearly 大得多。从分布的语域来看，almost 主要分布在小说语域中，而 nearly 则分布在报纸语域中较多。

（二）almost 和 nearly 的搭配词

搭配研究之父 Firth（1957）最早将“搭配”这一概念引入语言学领域，他将搭配视为一种独立于语法的词汇现象，并将其定义为“词项的习惯性结伴使用”，即词项间的习惯性共现，而搭配的重要性则在于“由词之结伴可知其词”^[4]。许多学者从不同的角度对搭配进行研究，所谓词语搭配其实并不仅仅是一种语法关系，更涉及语义和语用层面等。本研究采用了 Halliday(1964)对搭配的定义，即“体现词项某种显著的监控范围内组合关系的线性共现”，认为语料库的一套统计手段如节点词、跨距、搭配词等可以来测量这一“线性

共现”，即节点词与搭配词之间的相互吸引力(*mutual attraction*)^[5]。

本研究检索近义词 *almost* 和 *nearly* 的语境共现行，在近义词周围反复出现的词，便是 *almost* 和 *nearly* 的搭配词。语料库研究中通常把 MI 值等于或大于 3 的词作为显著搭配词。笔者在 BNC 语料库中以 *almost* 和 *nearly* 为节点词，选取跨距为(-5, 5)，分别对它们词进行检索，找出这组同义词各自的搭配词的共现频率和 MI 值，如表 5 和表 6。

首先，我们以 *almost* 为节点词，跨距-5/+5 检索其搭配词及 MI 值(见表 5)，其左右侧各 5 词之内“频繁出现”的搭配词按 MI 值>3 的大小排列，分别为 CERTAINLY,ENTIRELY,IMPOSSIBLE,IMMEDIATELY,EXCLUSIVELY,COMPLETELY,INVARIABLY,IDENTICAL,INEVITABLE,INEVITABLY,DOUBLED,UNIVERSAL,UNIVERSALLY,WHOLLY,INVISIBLE.从以上检索结果我们不难看出 *almost* 多与带有“im”前缀的形容词搭配，例如“不可能”，“不变的”，“不可避免的”，“不普遍的”等带有否定色彩的词语。

表 5 *almost* 在 BNC 中与其搭配词的频数及 MI 值
(跨距-5/+5)

搭配词	共现频数	总频数	MI 值
CERTAINLY	1510	17964	4.75
ENTIRELY	605	6689	4.86
IMPOSSIBLE	497	6761	4.56
IMMEDIATELY	453	10013	3.86
EXCLUSIVELY	304	1664	5.87
COMPLETELY	211	8237	3.04
INVARIABLY	210	1521	5.47
IDENTICAL	187	2130	4.81
INEVITABLE	99	2682	3.56
INEVITABLY	99	3025	3.39
DOUBLED	96	1251	4.62
UNIVERSAL	85	2552	3.42
UNIVERSALLY	81	577	5.49
WHOLLY	75	2208	3.44
INVISIBLE	59	1188	4.97

接着，我们以 *nearly* 为节点词，跨距-5/+5 检索其搭配词及 MI 值(见表 6)，其左右侧各 5 词之内“频繁出现”的搭配词按 MI 值>3 的大小排列，分别是：YEARS, HALF, CENT, MILLION, AGO, HUNDRED, NEARLY, FINISHED, THOUSAND, KILLED, 50, FORTY, 40, 100, TWICE。通过观察我们发现，与 *nearly* 搭配的词多为数词和时间词，比如“一半”，“一百”，“百万”，“一千”等。

表 6 *nearly* 在 BNC 中与其搭配词的频数及 MI 值
(跨距-5/+5)

搭配词	共现频数	总频数	MI 值
YEARS	976	87531	3.29
HALF	442	29116	3.73
CENT	364	37163	3.10
MILLION	359	23707	3.73
AGO	257	19057	3.56
HUNDRED	180	18576	3.08
NEARLY	146	11007	3.54
FINISHED	123	8204	3.71
THOUSAND	123	10106	3.41
KILLED	94	8238	3.32
50	91	7457	3.42
FORTY	82	6486	3.47
40	78	6058	3.49
100	77	7939	3.08
TWICE	76	5988	3.47

(三) *almost* 和 *nearly* 的类联接和语义韵

一个类联接代表了一个类别的词语搭配，比如：PREP + N 是一个类联接，它代表一类搭配即 PREP + N 搭配。基于以上研究，我们继续分析上述词语的语境共现，归纳近义词的常见搭配形式。通过查询 *almost* 的索引行，我们发现该词主要有下述搭配：*almost certainly*, *almost entirely*, *almost impossible*, *almost immediately*, *almost exclusively*, *almost completely*, *almost invariably*, *almost identical* 等，我们不难发现在 BNC 语料库中，*almost* 大多与副词和形容词搭配，它们的类联接型式有 *adv+adv* 和 *adv+adj*。

相应的，通过查询 *nearly* 的索引行，我们发现该词主要有下述搭配：*nearly thirty years*, *nearly half the counties*, *23.5 per cent to nearly 30 per cent*, *nearly 10 million*, *nearly thirty years ago*, *nearly two hundred pound*, *finished with nearly 1,900*, *nearly a thousand pound*, *nearly killed myself then*, *nearly 50 applicants*, *nearly forty five in January* 等，上述搭配词多为 *adv+num* 的结构。

某些词因循反复地与节点词在文本中共现，节点词就被传染上相关的语义特点，使整个跨距内弥漫着一种特殊的语义氛围。这种特殊的语义氛围便是语义韵。节点词即研究者语料库中检查其搭配行为的关键词。下面我们以前 *almost* 为节点词，跨距为 (0, 5)，MI 值>3 为检索条件，分析相关索引行。

(1) The sexual act, through a whim of nature's, has been made almost impossible to carry out.

(2) Sadly, the easiest way to capture a baby orang is almost invariably to kill the mother.

(3) It's almost equally inevitable that they will go down in the low income bracket

(4) the number of cases transmitted through heterosexual sex has almost doubled.

以上是 almost 与以“un”为前缀的形容词和副词的索引行,很容易发现这些句子都带有消极的语义韵。至于其他的形容词和副词,则要分不同的语域进行讨论,下面以 almost 的显著搭配词 certainly 为例,浅谈 almost 和 certainly 共现时,在不同语域的语义韵。

(5) Well almost certainly it will be, I think.

(6) I think we're almost certainly guaranteed to get the SLO's back into the fold

(7) Cowman, who had five stitches in a face wound, said the driver would almost certainly have been killed.

(8) Life for Mr Winner will almost certainly continue as before with him mistakenly believing he is a figure of some note.

(9) Though almost certainly more tolerant than the norms observed by the police in deciding whether or not.

(10) If he is so intoxicated, he is almost certainly guilty of the offence of being either drunk and disorderly

(5), (6) 句来自于会议语料, (7), (8) 来自于杂志语域, (9), (10) 来自于法律文本。观察发现,来自于会议语料的索引行大多带有积极或者中性语义韵,而来自杂志或法律文本的索引行多与“kill”, “mistake”, “guilty”等带有消极语义色彩的词语相关,因此,当 almost 和 certainly 搭配时,在会议文本中多为积极的语义韵,在杂志和法律文本中多为消极的语义韵。

同样的方法,我们以 nearly 为节点词,跨距为 (0, 5), MI 值>3 为检索条件,分析相关索引行。

(11) so that they are now at their lowest level for nearly 25 years. While this is good news for millions of borrowers and mortgage.

(12) Nearly half the women questioned in the General Household Survey, published by the Office

(13) we used to before the event, you would nearly always have the counts done.

(14) There are still nearly three million trade unionists in

either pre or post-entry closed shops.

(15) At present the country imports nearly 90 per cent of its timber requirements.

(16) The windows had been closed for nearly three months and the place had a dusty smell that got into your throat.

分析观察以上索引行,不难发现 nearly 的搭配几乎都是带有中性语义韵,由此可知, nearly 与 almost 搭配的语义韵是不同的。

五、结论

通过对 BNC 语料库进行相关语料的检索,本文对 almost 和 nearly 这一对近义词作了进一步的区分。通过对其在不同语域中的词频分布差异进行统计,检索行语料库中所呈现的近义词的显著搭配词,从而揭示出这一对近义词的搭配及语义韵等语言特征。研究发现, almost 和 nearly 有如下差异,第一,总体而言, almost 的使用频率在各语域比 nearly 大得多。从分布的语域来看, almost 主要分布在小说语域中,而 nearly 则分布在报纸语域中较多。第二, almost 多与带有“im”前缀的形容词搭配,例如“不可能”,“不变的”,“不可避免的”,“不普遍的”等带有否定色彩的词语。nearly 搭配的词多为数词和时间词,比如“一半”,“一百”,“百万”,“一千”等。第三, almost 大多与副词和形容词搭配,它们的类联接型式有 adv+adv 和 adv+adj, nearly 多为 adv+num 的结构。最后, almost 带有消极语义韵,而 nearly 多为中性语义韵。

[参考文献]

[1]陆军.基于语料库的学习者英语近义词搭配行为与语义韵研究[J].现代外语,2010(3):11.DOI:CNKI:SUN:XDWY.0.2010-03-013.

[2]吴小芳,肖坤学.基于语料库的英语动词近义词辨析的认知语言学研究[J].外语研究,2016,33(5):6.

[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.2014(5):45-47

[4]Firth, J. 1957. Papers in Linguistics. [M]London: Oxford University Press.

[5]Halliday M A K, McIntosh A, Stevens P.1964.The Linguistic Sciences and Language Teaching. London: Longman

作者简介:

易倩文(1995.6)女,汉族,湖北荆州,硕士,助教,认知语言学,英语教学