基于语料库的近义词辨析
——以名词 job， career， occupation 为例

梅家家
（天津外国语大学）
DOI：10.65285/JYLLYJ.V2I6.016

摘　要：近义词辨析一直都是困扰英语学习者的一个重难点。本文基于语料库的研究方法，运用 COCA 语料库从词频分布、语域分布、搭配语、语义韵方面对 job，career 和 occupation 进行分析和对比。研究表明语料库在英语学习者学习近义词辨析的过程中发挥了重要的作用。本文旨在帮助英语学习者可以通过语料库更高效准确的掌握近义词，并对其未来的英语学习提供新的更有效的方法。
关键词：COCA 语料库；近义词；语域；搭配；语义韵

English Synonym Differentiation Based on Corpus: Taking the Nouns job, career, occupation as an Example

Jiajia Mei
Tianjin Foreign Studies University

Abstract: Synonym differentiation has always been a major difficulty for English learners. Adopting the corpus-based research method, this paper uses COCA to analyze and compare job, career and occupation from the aspects of word frequency distribution, register distribution, collocations and semantic prosody. The study shows that corpora play an important role in English learners' synonym learning. This paper intends to help English learners master synonyms more efficiently and accurately with the help of corpora and provide new and more effective approaches for their subsequent English learning.
Keywords: COCA; synonyms; register; collocation; semantic prosody

前言：

英语词汇在英语学习中起着十分重要的作用，大多数教师都认为词汇是学习英语的基础，然而在学习词汇的过程中，英语近义词是一个重大难点。贺晓东（2003）约为60%，并通过词义和用法等方面的相同或相近组成了庞大且复杂的英语词汇体系[5]。因此在学习英语的过程中，近义词辨析也成了一个必须引起重视的问题。传统近义词辨析的方法大多是依赖词典的释义或者是例句，抑或个人直觉与语感。这些方法虽然有一定的作用，但是在实际操作过程中却有一定的难度，而且在某些情况下并不准确，所达到的效果也并不理想。

语料库是指按照一定的语言学原则，运用随机抽样的方法收集自然出现的连续的原文本或话语片段而建成的具有一定容量的大型电子文库，具有容量大，语料真实，检索快捷的独特优势[8]。本文以 COCA 语料库为基础研究名词 job， career， occupation 在词频、语域、搭配、历时分布和语义韵方面的异同，以便学习者更好的使用语料库，提高自己的词汇运用和辨析英语近义词的能力。

一、文献综述与理论基础

（一）理论概述

英国语言学家 Sinclair（1996）在其著作《寻找意义单位》中首次提出了“意义扩展单位”的概念。扩展意义单位也称为“词项”，它由四个部分组成：搭配、类联接、语义偏好和语义韵。他认为词语搭配指的是两个或两个以上惯常出现的词的共同出现，无独有偶。类联结是在搭配之后寻找意义扩展单位的下一步，是语法选择的共同出现。搭配词强调节点词与个体词汇之间的关系，类联接是指语法类别，即节点词与其他词共现的语法模式。语义倾向是指搭配词的语义特征，语义韵则被认为是找到意义扩展单位的最后一步。普遍认为，当找到语义韵时，词项的边界就接近了，语义韵可分为三类：消极、积极和中性语义韵。语义韵对词义的探索和研究有着不可或缺的作用。

（二）相关研究

英语词汇量众多，近义词对英语学习者来说是一个难题，也是词汇学习的重点。已经有不少专家学者意识到语料库对英语学习的重要性。任培红（2008）通过辨析 ordinary 和 common 两个形容词，发现传统的英语教学教学方法存在诸多问题，教师常常依靠词典释义，脱离具体语境，不能对近义词各项进行有效的辨析[10]；陆军（2010）在其文章中告诫众人：要在教学大纲和教材编写中重视词语搭配，重视基于语料库证据的近义词教学，英语中存在大量近义词，他们在搭配行为和语义韵上存在不同程度的差异，学会使用语料库对于增进近义词学习、掌握近义词意识、掌握近义词用法和提高近义词使用能力具有实用价值[9]。运用语料库进行英语近义词学习越来越需要加快普及和应用，可以有效提高非母语者学习英语的效率。

二、研究问题

本文旨在探讨和回答以下研究问题：

1) job， career， occupation 在词频分布上有何差异？

2) job，career，occupation 在不同语域分布上有何不同？

3) job， career， occupation 在搭配和语义韵方面有何差异？

三、研究步骤

（一）研究对象

对于英语学习者来说，英语近义词容易误用和混淆。“job， career， occupation”这三个英文单词在《柯林斯英汉大辞典》中都有“工作，职业”的含义，英语学习者容易对这三个词乱用。除此之外，有些人认为可以通过单词的英文释义进行辨析和理解，job， career， occupation 这三个英文单词的英文含义如表 1 所示。通过对其含义的解读，发现即使通过其英文含义进行辨析，也不能十分准确地区分，仍然存在似是而非的感觉，由此可知，无论是分析一个英文单词的直接中文意义，还是解析其英文释义，都不能帮助英语学习者高效准确地区分英语近义词。因此，本文以“job， career， occupation”为研究对象运用语料库进行分析，以便帮助英语学习者更好地理解和区分其含义。

表 1 job， career， occupation 在《柯林斯大辞典》中的英英释义

词项	英英释义
job	A job is the work that someone does to earn money.
career	A career is the job or profession that someone does for a long period of their life.
occupation	Your occupation is your job or profession.

（二）研究工具

本文使用 COCA 语料库作为研究工具，美国当代英语语料库 (COCA) 是由美国杨伯翰大学的 Mark Davies 教授开发的最新美国当代英语语料库，拥有 10 亿词汇量，是目前全球最大的英语平衡语料库。此外，COCA 具有很强的时效性，语料每年至少更新两万次，内容涉及近几十年来美国博客、网站、电视电影、口语、小说、杂志、报纸和学术期刊等八大类别的语料 [7]。它的数据可靠，凭借大规模和多样化语料，数据具有高度代表性，可以反映真实语言使用情况，避免因样本局限而导致研究偏差。

（三）研究方法

本文使用 COCA 美国当代英语语料库的在线检索功能和数据分析功能对 job， career， occupation 这三个英语名词进行分析。运用了该语料库的在线检索功能从词频，语域，历时分布，搭配以及语义韵方面解析和研究，从不同角度为英语学习者区分以及辨析 job， career， occupation 这三个近义词，可以帮助英语学习者便捷，高效，准确地掌握其含义。

四、研究结果与讨论

（一）词频分布差异

Aijmer K. 和 Altenberg B.（1991）认为，在一定程度上可以通过词汇的频率分布来展现语言的概率属性 [1]。本文首先通过语料库的 list 板块对 job，career，occupation 三个单词分别进行检索，以得出它们的词频，如表 2 所示。其中 job 出现的总词频数为 312985，career 出现的总词频数为 107773，而 occupation 出现的总词频数仅为 14177。由此可知，job 的词频数明显高于 career 和 occupation 的词频数。这说明在这一组近义词当中，job 的使用更为频繁，使用范围也会比另外两个单词更加广泛。

表 2 频数和频率

词项	频数	频率（每百万词）
job	312985	315.18
career	107773	108.53
occupation	14177	14.28

（二）语域分布差异

语域是指人们在实际的语言活动中，出于交际的需要或因其从事的职业和兴趣不一样而产生的一种演变体，体现为语言的不同语体风格，所以语义对于近义词辨析具有一定的意义，再辨析近义词时，结合语句可以直接的观察到他们的区别 [2]。

通过在 COCA 语料库中的 chart 板块分别输入 job，career 和 occupation，可以得到表 3——这一组词在 COCA 语料库中不同语域的词频分布情况。从表 3 可以看出，无论是在博客，网站，电视 / 电影，口语，小说，杂志，报纸以及学术期刊中，job 一词的频率都是最高的，career 一词仅次于其后，频率最少的则是 occupation 这一词。其中 job 频率最高的语域则是电视 / 电影这一类，说明它通常用于电视节目或者电影等娱乐产物中。而 career 经常用于的语域则是报纸当中，说明 career 一词偏向于用在正式文体当中。Occupation 一词出现频次最多的语域是学术期刊，出现频次最少的语域是电视 / 电影，也就是说其多用于学术以及正式文体当中。由此可见，通过分析近义词在不同语域中的优先用词原则，以便用语更加地道。

表 3 job， career， occupation 在 COCA 语料库中的词频及语域分布

近义词	总词频（每百万词）	博客	网站	电视 / 电影	口语	小说	杂志	新闻	学术文章
Job	315.18	368.38	344.87	424.28	365.59	241.97	255.86	363.74	142.98
career	108.53	114.16	112.77	52.63	89.58	42.36	134.79	214.28	108.06
occupation	14.28	12.25	16.46	4.36	12.49	5.98	14.82	19.02	29.47

（三）搭配差异

Firth(1957) 曾经说过一个单词的完整意义总是和上下文分不开的，我们想要了解一个单词，必须要了解它周围经常出现的单词，也就是说要了解这个单词的搭配 [3]。本文通过对 COCA 语料库中 Collocates 检索功能板块对 job，career 和 occupation 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，说明其与目标词的关联性越强，搭配强度越大。根据 Krishnamurthy(1987)，使用语料库检索的时候，在节点词的左右各 5 个词范围内对出现的搭配词进行观察，则具有研究意义。前后超过五个词即不具有研究意义 [4]。通过节点词和搭配词的 MI 值对词语搭配强度进行比对分析。MI 值 (Mutual Information Score) 即相互信息值，一般当 MI 值大于 3 时，则认为该搭配具有显著的搭配力，本文基于 COCA 语料库，将跨距设为 –3–+3，按照频数排序，并设定最小的 MI 值为 3，从高到低选取了 job， career， occupation 前 20 个搭配频数最高的词进行比对分析。

如表 4 所示，通过对 COCA 语料库的检索，前后与 job 搭配的词性有动词、形容词、名词。其中，动词出现 7 次，占 35%，它们分别是 doing， done， quit， lose， offered， finish， losing；形容词出现 4 次，占 20%，它们分别是 good， full–time， excellent， part–time；名词出现 9 次，占 45%，它们分别是 creation， growth， satisfaction， training， description， interview， opportunities， losses， seekers. 由此可见，job 一词通常与名词进行搭配，且搭配度最强的名词是 seekers。通过对 COCA 语料库进行检索，前后与 career 搭配的词性有动词，形容词，名词和介词，动词出现 4 次，占 20%，分别是 began， pursue， spent， ended；形容词出现 7 次，占 35%，分别是 his， political， early， entire， professional， successful， academic；名词出现 8 次，占 40%，分别是 college， path， development， choice， goals， opportunities， acting ，teaching；介词出现 1 次，占 5%，是 throughout，频次为 1600，说明 career 经常与介词 throughout 搭配使用。由此可知，career 经常与名词搭配使用，在形容词中，搭配度最强的是 successful，而在名词中，搭配度最强的是 path。通过对 COCA 语料库进行检索，前后与 occupation 搭配的词性有形容词，名词，介词，动词。形容词分别是 Israeli，military， foreign， German，west， Japanese，共 6 个，占 30%；名词分别是 Iraq， U.S.，Israel， forces， LBS， Soviet， invasion， age， education，Nazi， Kuwait，共 11 个，占 55%；介词有 during 和 under，占 10%；动词有 end，占 5%。通过对表中数据分析，发现 occupation 同样经常和名词进行搭配，且多和表示国家的名词进行搭配。

表 4 job， career， occupation 的高频搭配词

序号	搭配词	频次	MI	搭配词	频次	MI	搭配词	频次	MI
1	good	18545	3.12	his	27284	3.48	Israeli	672	7.69
2	doing	10852	3.64	began	2245	4.23	during	609	4.07
3	done	10769	4.00	political	1660	3.19	under	507	3.81
4	quit	2731	5.40	throughout	1600	4.77	Iraq	437	5.67
5	creation	2 628	5.20	early	1522	3.15	military	400	4.69
6	growth	2096	3.34	entire	1409	4.15	end	318	3.17
7	lose	2043	3.45	college	1357	3.39	U.S.	274	3.31
8	satisfaction	1874	5.53	path	1147	4.87	Israel	258	5.02
9	training	1862	3.14	development	1071	3.29	forces	252	5.01
10	description	1743	4.81	professional	1044	4.24	LBS	210	8.95
11	full–time	1641	5.87	successful	1043	4.36	Soviet	199	5.58
12	interview	1416	3.22	pursue	770	5.79	invasion	196	7.02
13	excellent	1268	3.84	choice	707	3.21	foreign	183	4.26
14	offered	1251	3.19	spent	692	3.05	age	183	3.40
15	finish	1238	3.57	ended	690	4.02	education	173	3.13
16	opportunities	1205	3.67	academic	668	4.14	German	171	5.40
17	losing	998	3.21	goals	625	4.13	Nazi	169	7.46
18	part–time	987	5.84	acting	610	4.41	Kuwait	167	7.07
19	losses	971	4.42	opportunities	547	4.07	west	164	3.72
20	seekers	818	7.10	teaching	487	3.15	Japanese	131	4.95

（四）语义韵分析

Sinclair 通过借用 Firth 的“音韵 (prosody)”一词而提出了语义韵 (Semantic Prosody) 这一概念。研究表明，一些节点词总是与一类具有相同或相似特点的搭配词在文本中反复共现，久而久之，节点词也被传染了这些搭配词的语义特点，这就是语义韵。Stubbs 基于语义韵的概念，将其分为三类：积极 (positive) 语义韵、中性 (neutral) 语义韵和消极 (negative) 语义韵。

通过表 4 可以看出，job 多呈现出中性语义韵的特点。job 本身是中性词汇，不带有固定的感情色彩。当它与积极的词汇搭配时，可以呈现出积极语义韵，例如表 4 中的 good，excellent 这些词，当它与负面的词汇或者语境结合时，就会展现出消极语义韵的特点，例如 lose a job 和 quit a job。career 一词更加侧重社会地位与个人成就，那么它的积极倾向更强，与它搭配的形容词如 successful，professional，academic 具有积极的感情色彩；最后是 occupation，这个词多用于学术期刊或者报纸中，相比较于前两个单词，occupation 更加学术化和官方化，更倾向于中性语义韵。

结语：

本文基于 COCA 美国当代英语语料库，对 job， career 和 occupation 这组近义词进行了对比和分析。通过语料库对其检索和研究，分析了这一组词在词频、语域分布、搭配词和语义韵的差异，并得出以下结论：1)job 的使用频率远高于 career 和 occupation，occupation 是使用频率最少的。2) job 多用于非正式文体，如口语或者电视电影中，career 和 occupation 多用于报纸或者学术期刊的正式文体当中。3)job 和 occupation 呈现出中性语义韵，而 career 更倾向于积极语义韵。

参考文献：

[1]AIJMER K, ALTENBERG B, SVARTVIK J. English corpus linguistics: studies in honour of Jan Svartvik[M]. London: Longman, 1991.

[2]HALLIDAY M A K, MCINTOSH A, STREVENS P. The Linguistic Sciences and Language Teaching[M]. London: Longman, 1964.

[3]FIRTH J R. Papers in Linguistics[M]. London: Oxford University Press, 1957.

[4]KRISHNAMURTHY R. The process of compilation// Sinclair J M. Looking up: An Account of the Conbuild Project in Lexical Computing[M]. London: Collins ELT, 1987.

[5] 贺晓东 . 英汉双解英语同义词词典 [M]. 北京 : 商务印书馆国际有限公司 ,2003.

[6] 卫乃兴 . 词语搭配的界定与研究体系 [M]. 上海 : 上海交通大学出版社 ,2002.

[7] 汪兴富 ,MarkDavies, 刘国辉 . 美国当代英语语料库 (COCA)——英语教学与研究的良好平台 [J]. 外语电化教学 ,2008(05):27-33.

[8] 杨惠中 . 语料库语言学导论 [M]. 上海 : 上海外语教育出版社 ,2002.

[9] 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(03):276-286+329-330.

[10] 任培红 . 近义词的教与学 : 以 Common 和 Ordinary 为例 [J]. 解放军外国语学院学报 ,2008(04):57-60.

作者简介：梅家家（2001 年 -）、女、河南驻马店人、天津外国语大学、硕士研究生、研究方向：英语语言文学。