

语料库数据驱动下的英语四级词汇教学研究

王艾艾

武警工程大学基础部, 陕西 西安

收稿日期: 2024年6月15日; 录用日期: 2024年7月18日; 发布日期: 2024年7月25日

摘要

本研究运用语料库语言学技术, 辅以人工词汇语境分析, 定量与定性研究相结合, 基于教材语料库、四级真题语料库和大学英语四级核心词汇, 从研究目标、研究内容和研究步骤三个方面, 阐述如何在语料库数据驱动下, 创新大学英语词汇教学模式, 以期适应现阶段院校大学英语大纲和人才培养方案修订等教学改革新举措, 满足大学英语四级教学需求, 发挥学员主体作用, 提高学习效率, 实现教学目标。

关键词

语料库, 数据驱动, 英语四级, 词汇教学

A Study on CET-4 Vocabulary Teaching and Learning Driven by Corpus Data

Aiai Wang

Foundation Department, Engineering University of PAP, Xi'an Shaanxi

Received: Jun. 15th, 2024; accepted: Jul. 18th, 2024; published: Jul. 25th, 2024

Abstract

This paper combines both quantitative and qualitative research based on corpus linguistics technology and context analysis of vocabulary. Through self-built textbook corpus, CET-4 exam paper corpus, and core vocabulary of CET 4, three aspects are discussed to innovate the teaching and learning mode of college English vocabulary: research objectives, research content, and research steps. Driven by corpus data, study of the new teaching and learning mode aims to adapt to new teaching reform measures such as revising the current college English syllabus and talent training programs. In this case, the teaching requirements of CET 4 can be fulfilled with students playing the major role, to finally improve learning efficiency and achieve teaching objectives.

Keywords

Corpus, Data-Driven, CET-4, Vocabulary Teaching and Learning

Copyright © 2024 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

目前,各高校都在积极探索大学英语教学新模式,从课时量、内容设置、思政融入等多个角度进行教学改革,解决当前大学英语教学存在的问题。以笔者所在高校为例,在大学英语词汇教学方面,呈现出不少问题:一是部分老师对大学英语四级词汇教学的重要性认识不足,满足于传统教学模式,缺乏创新的积极性。认为词汇的积累是学生自己的事,大学英语课不需要花大力气去思考词汇教学,习惯于采用例句翻译和课堂听写的方式完成对词汇的讲解和检查。二是部分老师词汇教学能力不强,相关教学能力和认知度仍需深化。没有对当前院校教学改革、人才培养方案修订和课程教学时间和内容调整进行深入思考,对最新的教学理论和方法缺乏研究,没有将前沿的理论和应用于教学实践。三是部分老师在词汇教学中未能把握重点,没有结合学生实际进行设计和调整,所讲授的例句有的是基于经验的编写,有的是网络例句的照搬照抄,没有进行精心筛选,甚至出现讲解偏差和错误,无法达到理想的教学效果。

上述问题的出现是对词汇知识的三个方面的忽视:1) 口语或书面中遇到该词的概率;2) 该词的句法行为;3) 该词与其他词构成的关联网络和语用功能[1]。仅仅参考字典释义,通过简单的例句练习,无法对语言的本质性差异进行深入研究,还需通过分析海量相关语料来分析词汇之间更全面更科学的特征差异。因此,本研究将专注于语料库数据驱动下的词汇教学研究。

2. 研究目标

基于目前大学英语教学,尤其是词汇教学中存在的问题,结合同类院校教学实际,我们主要围绕以下三个目标开展相关研究。

2.1. 更新语料库建设成果

作为一种语言教学手段,语料库应用于课堂教学有其独特优势:呈现语言知识、吸引学生注意力、帮助学生记忆、开发学生学习智慧、灵活设计练习[2]。笔者所在研究团队前期已建设完成各类军事英语语料库以及教材语料库,为课堂教学和学员自学提供了重要的真实的语料,有效提高了课堂效率。本研究拟对语料库建设进行更新,建设完成1989年至2023年历年四级真题语料库,为四级教学尤其是四级词汇教学提供重要素材和考核依据。

2.2. 深入研究语料库视域下词汇教学的创新性实践效果

本研究运用语料库语言学技术,辅以人工词汇语境分析,定量与定性研究相结合,基于教材语料库、四级真题语料库和大学英语四级核心词汇,从创设词汇学习语境、制作词汇教学微本等方面,创新大学英语词汇教学模式,以期适应现阶段院校大学英语大纲和人才培养方案修订等教学改革新举措,满足大学英语四级教学需求,发挥学员主体作用,提高学习效率,实现教学目标。

2.3. 寻求基于各类语料库开展教学实践的推广应用

语料库应用于大学英语教学中,学员可以在真实语境下提高对词汇表达的理解和掌握,在探索学习中发现以往被忽略的语言事实和规律,在反复的观察和分析中发现自己在学习中的问题并及时改正,这些都极大地提高了学员的学习效率和学习兴趣,培养了自主学习能力和跨文化交际意识[3]。本研究在为同类院校提供一定的借鉴和帮助的同时,可开展校际间合作研究,力求在更大范围内推广应用。

3. 研究内容

针对目前院校教学改革和人才培养方案修订实际,以四级词汇教学这一重要教学模块作为研究对象,研究构建大学英语四级真题语料库(College English Test Band Four Corpus,简称 CET4C),利用语料库语言学软件和技术,采用定性和定量研究相结合的方式,为教学创设词汇学习语境、制作词汇教学微本、提供学员自学工具。具体研究内容如下:

3.1. 大学英语词汇教学研究情况回顾与分析

将文献资料区分不同类别进行总结分析,如:基于语料库分析的大学英语词汇教学研究与定性研究为主导的大学英语词汇教学研究;基于大规模综合型语料库,以语料库驱动学习的方法,引导学生主动发现词汇搭配特点、语法特点、不同语境下的语言规律,提升词汇学习效率和能力的研究;从语料库内涵出发,分析学生英语词汇学习中存在的困难和问题,介绍语料库运用于词汇教学中的具体方法和策略的研究;根据教学实际需要,自建相关语料库,将例句融入教学,凸显词汇语境,引导学生通过探索式、发现式、体验式学习模式,获得更好的词汇学习效果的研究等。对比分析后进行归纳总结,找出前人研究的特点和缺口,为本研究内容的确立提供更加具体和详实的参考依据。

3.2. 语料库建设和运用研究

在学习掌握语料库语言学理论知识和熟练运用语料库相关检索软件的前提下,建设 CET4C。按照考试时间、考试题型(写作、听力理解、选词填空、完形填空、快速阅读、仔细阅读、语法与结构、翻译等)和套题数等对 CET4C 按照不同类别分别存储命名,便于检索时快速定位文本来源和题型,开展专项内容分析解读,同时,对每种题型出现的频次(篇数)进行统计,有助于研究结果更加客观和准确。利用语料库检索软件,对四级核心词汇在教材和四级真题中的应用情况进行检索分析,在课堂教学中为学员展示词汇在真实语料中的语境,通过分析词汇在不同语料库中的出现频次(frequency)、搭配特点(collocation)、类联结(colligation)、语义倾向(semantic prosody)、语义特点等,掌握词汇用法,基于客观数据,结合对词汇内容的把握和理解,开展定量与定性相结合的研究。

3.3. 相关教学活动探索和研究

一是开展语料库语言学讲座辅导及文献阅读与分享活动,普及相关理论知识、软件操作方法,设计教学活动;二是发挥教师主导作用,制作多类语料库教学微本,用于课堂教学实践,提高课堂教学效率;三是开展基于语料库的丰富多彩的趣味词汇竞赛,提升学员学习兴趣和语料库软件操作能力;四是开发建设多个门类的英语单语、英汉双语语料库,贴近学员学习实际以及未来岗位任职需求,对教材内容进行合理有效补充。

通过以上内容的研究,满足院校教学改革和人才培养方案修订背景下对大学英语教学提出的新要求,使大学英语教学育人功能更加完备,为一线教员提供操作性强、教学效率更高的四级词汇教学路径;同时创新性提出运用动态更新的英语语料库辅助四级词汇课堂教学和自学,对传统四级词汇教学模式进行

革新,探索大学英语四级词汇教学新路径,也为其他同类院校大学英语四级教学提供一定的借鉴。

4. 研究方案

4.1. 研究分工及文献阅读梳理

首先,对研究进行整体规划和过程监督。制定研究的整体实施方案,确定具体的分工和任务安排,对研究各个步骤的落实进行监督管理,确保研究过程有序,研究结果有效;针对语料库语言学相关理论知识和操作技能,开展有针对性地培训辅导,掌握语料库语言学基本概念和检索方法,为建设语料库,对语料库检索行进行分析、解读奠定基础。

其次,对收集语料进行格式转化、去除噪声信息、分类存储,按照指定格式分类存储,建设完成教材语料库以及报告撰写。对语料库检索结果进行导出分析,归纳总结四级核心词汇在教材和四级真题中的分布特点,撰写分析报告。利用学校图书馆电子阅览室,开展国内外相关研究的检索、分析和研究,从宏观上把握当前的研究方向、研究重点和研究缺口,为本研究提供参考和借鉴。

4.2. 四级真题语料库创建

从文献检索来看,许多研究仍以定性研究为主,缺乏有力的数据支撑,研究的可信度不足,研究结果存在不同程度的争议。而随着语料库语言学的发展,语料库以其科学、客观、直观的定量研究为研究人员提供了更多真实可靠的数据来源,使得研究结果更具参考价值,可信度更高[4]。因此,基于语料库数据对教材进行研究已变得越来越普遍。

团队拟自建大学英语四级真题语料库 CET4C。语料的获取途径主要利用在线网络资源,如百度文库、可可英语网、新东方四六级网、文都网等进行资料下载和格式转化,将相关纸质试题进行扫描,转化格式,同时使用 Ediplus 进行降噪预处理,以语料库软件可识别的 txt 文本格式进行保存。

CET4C 包含 1989~2023 年大学英语四级真题共 122 套,内容按照历年四级考试题型分为写作、听力理解、选词填空(完形填空)、快速阅读、仔细阅读、语法与结构、翻译等七个部分,每部分均按照考试不同年份、套题数和题型类别进行存储命名,便于进行检索时能够快速定位文本来源和类别,开展专项内容分析解读。

为便于语料库信息检索,CET4C 拟将涉及的试题材料按照独立文件进行存放,文件以“年份 + 月份 + 套题数 + 具体试题类型”方式命名。例如,2012 年 6 月第一套题听力篇章命名为“2012-6-1-listening comprehension”,2023 年 12 月第三套题深度阅读命名为“2023-12-3-in-depth reading”。同时,考虑到语料库的快捷检索,每个年份所有语料的还要进行统一存放,命名形式为“年份”,因此,各年份的命名分别为“1989 年”、“1990 年”、“1991 年”、“1992 年”等,即 CET4C 的 35 个子库,涉及文本共 943 篇,其中写作文本 76 篇,听力文本 113 篇,选词填空(完形填空)文本 101 篇,快速阅读文本 74 篇,仔细阅读文本 355 篇,翻译文本 88 篇,语法和词汇文本 36 篇。

4.3. 词汇检索和语境分析

在学习掌握语料库语言学知识以及掌握语料库检索软件基本操作技能的基础上,将四级核心词汇关键词在教材和四级真题语料库中进行检索并对结果进行系统分析,即运用定性和定量研究相结合的方式开展研究。定量研究工具为语料库检索专用软件 AnConc3.2.1,运用其语境共现行(concordance),对词频(frequency)、词汇搭配(collocation)、类联结特点(colligation)、语义倾向(semantic prosody)以及不同语境下的语义特点进行分析解读,最后归纳总结。

以《新视野大学英语读写教程第四册》第七单元单词“rate”为例,它在课文中的语境为“The death

of an angel of animal rights activism does not *rate* with that of a drugged-out rock star”。*rate* 在该句中为不及物动词，做谓语，与介词“with”构成固定搭配，表示“受青睐，受器重”。为了解 *rate* 在课本中更多的语境和含义，我们将其输入教材(《新视野大学英语读写教程第三版》全四册)语料库中进行检索。

从检索结果看：*rate* 在教材中出现频次为 5 次，其中做主语 3 次(与名词或形容词构成固定搭配)，宾语 1 次，谓语 1 次(固定搭配)，由于频次较少，无规律性特点。与 *rate* 共现的主题词汇包括 *bankruptcies*, *unemployment*, *economic growth*, *injury*, *fatality*, *turnover*, *death*，既包含褒义词和贬义词，也有中性词汇，语义倾向不明确，与检索行较少有关，这是 *rate* 在教材中的语义韵特点。而 *rate* 在 5 种语境中的词义分别为(失业)率，(经济发展)速度，(死亡和严重受伤)率，(周转)率，受青睐、受器重。主要以名词形式出现，与不同词汇构成相应的固定表达。由此可见，不同语境下单词的含义往往不同。教师应结合具体的语境对单词的使用方法进行讲解，将教材语料库中所有语境进行分类呈现，易于学员把握该词汇在教材不同语境下的不同含义。

学员掌握 *rate* 在全四册教材中的使用语境和基本用法后，继续将其输入四级真题语料库 CET4C，在更大范围的语境中观察该词汇是否还有更多的延伸涵义。从检索结果看，*rate* 在 CET4C 共出现频次为 97 次，是四级大纲中的核心词汇。从题型分布看：*rate* 在阅读理解中出现 68 次，听力 21 次，语法与结构 4 次，写作 2 次，翻译 2 次。从语境分析看：*rate* 做主语 51 次(常与名词或形容词构成固定搭配)，状语 23 次，宾语 14 次，谓语 9 次。在真题语料库中，该词出现明显的频次差异，主语最多，其次为状语和宾语，谓语最少。因此，学员在使用时应注意其频次较高的主语和状语用法。与 *rate* 共现的主题词汇包括 *modern life*, *interest*, *heart*, *advertising*, *unemployment*, *population*, *education*, *return*, *online student*, *inflation*, *quality* 等，包含褒义、贬义及中性含义，与教材语料库检索结果一致。而 *rate* 在真题语料库不同语境下包含四种含义：与介词搭配表示“以……速度/速率”、“无论如何/不管怎样”，与名词或形容词构成固定搭配“(出生)率、(死亡)率、(离婚)率、(失业)率”等，做及物动词表示“评价/评估”，做不及物动词与 *with* 搭配构成固定搭配“*rate with*”。四级真题语料库中由于库容更大，语境更全面，包括了 *rate* 的所有常用词义，有助于深刻理解 *rate* 在不同语境下的多种含义，而语料库检索软件的快速检索功能也便于对词汇所有语境含义进行归纳总结。

在具体教学中，可根据分析结果，结合教材和四级考核内容，对核心词汇语境共现行进行分类整理，制作教学微本，用于课堂教学，减少大规模语料检索带来的时间不足问题。同时，将简单易操作的内容作为课前课后自学需要完成的作业，充分调动学员学习积极性，弥补课时减少带来的问题。

4.4. 相关教研活动的实施

一是开展语料库语言学讲座辅导，普及相关知识和操作方法，便于开展教学和相关研究。语料库语言学作为一个较为独立的研究方向，具有较强的系统性和专业性，所涉及知识和软件操作方法具有专业性、多样性和实用性等特点，对于没有任何基础的教师、学员或科研人员具有一定挑战性，短期内提升存在一定困难。因此，有必要在研究前期开展语料库语言学专业基础知识和基本软件操作技能的培训和辅导，并在研究全程进行指导和帮助，才能按时高效完成相关研究任务，取得理想的研究成果。

二是发挥教师主导作用，制作多类语料库教学微本，用于课堂教学实践，提高课堂教学效率。语料库以其内容丰富、真实和更新度高而具有教材或其它纸质材料不可替代的优势，但对于运用于具体课堂教学中的语料，需要考虑到教学环境和教学条件的局限性可能造成的课堂效率低下的问题。因此，根据教材内容、教学计划安排和学员具体学习情况，将语料库制作成适应课堂教学实际的教学微本很有必要且意义重大。

三是开展基于语料库的丰富多彩的趣味词汇竞赛，提升学员学习兴趣和语料库软件操作能力。将各

类教学微本进行整合, 组织开展各类趣味英语词汇竞赛, 如英语思政词汇比赛、英语四级高频词汇大赛、英语军事类词汇大赛、优秀影视作品词汇大赛、用英语讲好中国故事词汇比赛等, 充分调动学员的学习积极性, 潜移默化地提升教师和学员对语料库软件的操作能力和学员对各类词汇的理解和掌握能力。

四是以学员为主体, 结合学员的学习实际和未来岗位任职需求, 研究团队开发建设多个门类的英语单语、英汉双语语料库, 对教材内容进行有效补充。词汇教学作为大学英语教学的重要组成部分, 主要依托教材文本展开, 因此, 缺乏与时俱进、真实语境下的语言支撑, 课堂教学模式创新不足, 存在学用脱节问题。学员可以通过及时更新的语料库了解各国语言发展、历史、文化、经济、政治等最新信息, 提高对英语语言的认知, 为未来岗位任职打下基础, 还可以在真实语境下提高对各类词汇和表达的理解和掌握, 在探索学习中发现以往被忽略的语言事实和规律, 在反复的观察和分析中发现自己在学习中存在的问题并及时改正, 这些均能极大提高学员的学习效率和学习兴趣, 培养其自主学习能力和跨文化交际意识。因此, 语料库应用于教学可作为词汇教学创新的重要措施进行推广使用。

5. 结语

本研究基于语料库语言学理论和技术, 是数据驱动下的大学英语词汇教学创新和发展。我们将持续与其他同类院校大学英语教研团队进行合作交流, 在语料库语言学技术推动下, 不断更新教学类语料库, 拓宽研究思路和方向, 开发新的教学资源, 推出新的教学活动, 促进大学英语教学改革, 培养更多更优秀的适应国家建设和发展的复合型外语人才。

参考文献

- [1] 濮建忠. 英语词汇教学中的类联结、搭配及词块[J]. 外语教学与研究, 2003, 47(6): 438-445.
- [2] 赵志刚, 何安平. 报刊语料库在航海英语阅读教学中的应用[J]. 广东海洋大学学报, 2014, 34(5): 80-85.
- [3] 王艾艾, 王姣. 军事英语语料库在军事英语教学中的应用[J]. 武警工程大学学报, 2021, 37(1): 89-92.
- [4] 王艾艾. 基于语料库分析的军校大学英语读写教材思政功能研究[J]. 新教育时代, 2023, 12(5): 122.