

基于语料库的民办高校英语专业教学资源开发研究

陶思逸¹, 谭元昌²

¹重庆城市科技学院人文学院, 重庆

²重庆交通大学外国语学院, 重庆

收稿日期: 2024年7月24日; 录用日期: 2024年8月24日; 发布日期: 2024年9月4日

摘要

语料库是语言研究和学习的重要工具。对于民办高校来说, 如何利用这项工具培养应用型英语人才是语言专业发展的趋势和重点。针对民办高校普遍存在的英语专业学生英语语言应用能力培养教学资源不充分的特点, 本研究自建英语教学素材语料库, 通过量化分析英语新闻文本的方法, 高效提炼英语教学资源, 以此拓宽民办高校英语专业阅读能力培养路径, 对今后的民办高校英语教师的教学起到一定的启发。

关键词

语料库, 民办高校, 英语专业, 教学资源开发

Research on the Development of Educational Resources for English Majors in Private Colleges Based on Corpus

Siyi Tao¹, Yuen Cheong Tam²

¹School of Humanities, Chongqing Metropolitan College of Science and Technology, Chongqing

²School of Foreign Languages, Chongqing Jiaotong University, Chongqing

Received: Jul. 24th, 2024; accepted: Aug. 24th, 2024; published: Sep. 4th, 2024

Abstract

The corpus is an important tool for language research and learning. For private universities, utilizing this tool to cultivate application-oriented English talents is a key trend and focus on the development of language programs. Addressing the common issue of insufficient teaching resources for

developing English language application skills among English major students in private universities, this study constructs a self-built corpus of English teaching materials. By applying quantitative analysis methods to English news texts, this study efficiently extracts English teaching resources, thereby broadening the paths for developing reading skills in English majors at private universities. This research offers valuable insights for future English teaching practices in private universities.

Keywords

Corpus, Private Colleges, English Major, Development of Teaching Resources

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

随着网络媒介的快速发展,数智化教育是大势所趋。学生能通过网络接触到大量的信息,而其中的语言信息碎片化、模式化、独立化,使得学生所接纳的信息良莠不齐。李威(2020)指出,要利用网络信息进行语言学习需要将网络信息标准化,而语料库就是一个整理保存各种形式语言资源的载体。语料库资源体系是由各个相互作用、相互联系的语言库整合成的有机知识体系[1]。语料库研究对于高校一线教师如何利用数智化资源进行高效教学和改革有着重要的意义。如何充分利用语料库的特点开发优质的英语教学资源,成为了高校英语专业普遍面临的课题。

2. 文献综述

2.1. 民办高校英语专业学生就业形势与教学资源开发需求

目前外语专业需要通过进行一系列改革以应对就业竞争不断增大的现状。在此背景下,民办高校面临英语专业发展的瓶颈期,只有找准自身定位,其改革方案才能有的放矢。杨子莹(2018)针对民办高校的调查表明,2012、2013和2014级60名英语专业毕业生的就业情况,从事英语相关工作占30%,其中高考志愿被调剂到英语专业的占60%。数据在一定程度上反映了民办高校英语专业普遍存在生源专业基础薄弱、学习动机欠缺、英语应用能力不足等问题[2]。任安莉(2023)认为,“双减”政策实施后,教培业的英语人才需求大幅下降,给英语专业学生的未来发展严峻考验[3]。面对该局面,各行业对英语专业毕业生的应用能力提出了更高的要求。Smith, B. (2018)的研究表明,提升学生的实际应用能力,特别是语言运用能力,是提高就业竞争力的关键因素[4]。民办高校如何在传统教学资源基础上进一步挖掘和开发新的教学资源,以应对这一挑战,成为一个重要课题。

2.2. 语料库在语言学习中的应用与发展

近年来,计算机技术的迅速发展也推动了语料库语言学的发展,语言库研究的模式也有所变化。Wolfgang, T. (2005)指出,语料库研究几乎已成为所有语言研究的重要组成部分[5]。语料库中存储着数以亿计的语言数据集合,为外语学习提供了丰富的教学资源。田野,李继民,柳婷婷(2012)指出,在实际英语教学中能够对真实语言交际的各层面,包括词汇、语法、语篇等进行深入分析,对归纳英语的语用规律,引导学生习得自然、地道的英语具有重要作用[6]。语料库的实用性特点与民办高校的英语专业培养目标契合。例如,Zhang, W. (2021)指出,收集实际使用的英语文本,如我国官方媒体的英语新闻文本,

构建英语教育语料库, 并对其进行整理和分析, 可以提炼出丰富的英语教学资源[7]。

2.3. 语料库应用于外语教学的理论基础和研究现状

语料库语言学在外语教学中的应用逐渐受到重视。Sinclair, J. (1991)认为, 语料库语言学不仅仅是一个数据收集工具, 更是揭示语言规律的重要手段[8]。通过语料库, 可以真实反映语言使用的频率、搭配和模式, 有助于学生掌握实际语言运用中的细微差别。Bernardini, S. & Aston, G. (2004)指出, 数据驱动学习(Data Driven Learning, DDL)能够提高学习者的自主学习能力, 促进对语言规律的探索和理解[9]。李丽丽(2020)探讨了语料库在教学中的具体操作方法, 并提出了在不同教学阶段应用语料库的策略[10]。综上, 国内外学者对语料库在外语教学中的应用进行了广泛研究, 语料库在提高学生语言感知和实际应用能力方面具有显著优势。

2.4. 具体研究问题及解决方法

语料库语言学在外语教学中的应用逐渐受到重视。语料库不仅仅是一个数据收集工具, 更是揭示语言规律的重要手段。通过语料库, 可以真实反映语言使用的频率、搭配和模式, 有助于学生掌握实际语言运用中的细微差别。语料库的应用不仅能够提供丰富的语言输入, 还能帮助学生进行自主学习和探究, 提升语言敏感度和使用能力。数据驱动学习能够提高学习者的自主学习能力, 促进对语言规律的探索和理解。通过 DDL, 学生可以在教师的引导下使用语料库搜索和分析真实语言数据, 发现语言使用中的模式和规律, 从而提高语言实践能力。语料库在提高学生语言感知和实际应用能力方面具有显著优势。通过语料库, 学生可以接触到大量真实的语言材料, 增强对语言使用情境的理解。然而, 目前民办高校英语专业教学实施过程中对语料库的利用程度较低, 难以丰富教材以外的教学资源。因此, 本研究旨在解决以下问题, 如何通过自建语料库来丰富民办高校英语专业的教学资源提高学生的英语应用能力, 以增强民办高校英语专业学生的就业竞争力。

3. 研究方法

3.1. 语料库构建

本研究以《人民网英文版》(People's Daily Online)、《新华网英文版》(Xinhua News)、《中国日报英文版》(China Daily)等多个权威新闻网站作为数据来源, 收集时间范围为 2023 年 6 月到 2024 年 6 月, 涵盖一年的时间跨度, 以确保语料的时效性和连续性, 关键词为“Chongqing”(重庆)的英语新闻文本。为提升语料库的代表性, 本研究不仅收集新闻报道, 还包括社论、专栏文章、评论、特写报道等不同体裁的文本。进一步考虑不同语域(如新闻、经济、文化、科技等)的语言特点, 确保语料库涵盖广泛的主题和语言风格。抓取每篇英语新闻文本的标题、时间及正文内容, 排除重复的新闻文本后, 总共获得 2500 篇新闻文本。对得到的文本进行数据清理, 包括去除 HTML 标签及其他无关字符, 使用自然语言处理技术进行分词和词形还原, 以确保单词形式的一致性。为后续的定量分析提供更可靠的数据支持, 并提升研究结果的可靠性和普适性。最终形成了一个包含约 1,000,000 词的语料库。

3.2. 量化分析方法

量化分析作为一种实证研究方法, 旨在借助统计科学等手段将揭示隐含在大量语言数据中的语言规律。本研究将采取多种数据分析方法, 从语料库中提炼英语教学资源。在关键词分析中, 使用 TF-IDF (Term Frequency-Inverse Document Frequency)来衡量词汇的重要性, 确定重点词汇列表, 通过关键词分析, 可以为词汇教学提供数据支持, 帮助学生掌握高频词汇及其使用场景。通过依存句法分析工具(如 SpaCy)分析文本, 提取文本中常见词汇搭配, 记录高频出现的搭配。常见搭配分析有助于把握新闻文本中的常见搭

配，为语法教学提供真实的语言材料，增强学生对词汇和短语的实际运用能力。使用 LDA (Latent Dirichlet Allocation)等主题模型方法识别文本中的主题。通过 Count Vectorizer 将文本转化为词袋模型，应用 LDA 模型识别文本中的潜在主题，提取出新闻文本中的主要话题，从而帮助设计主题明确的教学资源。采用语篇分析方法，分析文本的宏观结构和信息组织方式，包括主题的展开、信息的衔接和连贯性等。语篇分析能够帮助学生理解和掌握不同类型文本的结构特征，提高对长篇文本的理解和处理能力。通过这些分析方法，本研究将全面挖掘语料库中的语言信息，为民办高校英语专业的教学资源开发提供数据支持和理论依据。

3.3. 教学资源开发路径

本研究以定量分析方法的结果为基础开发教学资源，在词汇教学方面，根据词频分析结果，设计出包含高权重关键词的词汇表。关键词反映了实际使用中的重要词汇，有助于学生在日常交流中灵活运用。在阅读理解练习方面，选择具有代表性的新闻段落，设计阅读理解问题，选择典型段落并设计阅读理解问题，培养学生的阅读技巧。在英语写作方面，根据主题模型分析和语篇分析结果，确定新闻文本中的主要话题，深入把握文本的段落结构和逻辑关系，并据此设计写作任务。学生通过撰写与新闻主题相关的作文，从语篇分析结果中了解该类文本的主要信息组织方式，加强文本的构建逻辑，从而提升写作能力，同时深化对所学内容的理解。在英语口语方面，利用新闻文本中的事件设计讨论话题，鼓励学生在课堂上进行口语表达和讨论，增强语言运用能力和表达自信。以上从多个角度将定量分析结果转化为具体的教学资源，通过真实语言材料的应用，全面提高学生的英语应用能力，实现语言应用能力的提升。

4. 从语料定量分析结果开发教学资源

4.1. 文本关键词分析

通过对语料库进行关键词统计分析，本研究识别出了在实际语言使用中最常见和重要的词汇。该关键词表不仅反映了文本的核心内容，还为英语教学提供了数据支持，帮助设计具体的教学资源。以下表 1 是关键词及其 TF-IDF 得分：

Table 1. Top 50 keywords and their weights
表 1. 前 50 位关键词及其权重

Keyword	TF-IDF Score	Keyword	TF-IDF Score
Chongqing	21.31	company	8.71
development	20.59	vehicles	8.70
tourism	15.81	products	8.70
cooperation	14.57	production	8.62
countries	13.17	holiday	8.45
trade	13.11	southwest	8.38
yuan	12.43	energy	8.22
market	12.08	equipment	8.22
growth	11.40	innovation	8.18
city	11.23	consumption	8.16
tourists	10.28	water	8.03
industry	10.27	conference	8.00

续表

regions	10.09	China Europe	7.92
municipality	10.02	night	7.92
economy	9.71	efforts	7.90
railway	9.61	manufacturing	7.84
technology	9.50	corridor	7.83
authorities	9.41	services	7.78
travel	9.27	drone	7.74
world	9.23	village	7.69
train	9.13	enterprises	7.54
road	9.00	investment	7.52
freight	8.83	construction	7.51
festival	8.73	transportation	7.47

1) **词汇教学**: 设计包含高频关键词的词汇表, 并围绕关键词拓展词表, 且为关键词提供应用于该类新闻文本时的定义、例句和相关练习, 帮助学生掌握实际使用中的常见词汇。

2) **阅读理解练习**: 选择包含高频关键词的新闻段落, 设计阅读理解问题, 如段落大意、词汇解释和内容细节, 帮助学生在真实语境中理解和应用该词汇。

3) **英语写作**: 根据高频关键词的主题设计写作任务, 提升学生的写作能力, 如撰写关于重庆旅游业发展的文章, 要求使用“tourism”、“development”等关键词。

4) **英语口语**: 利用高频关键词设计讨论话题, 鼓励学生在课堂上进行讨论和交流, 如讨论重庆的经济发展和贸易合作, 使用“trade”、“cooperation”、“economy”等关键词。

通过上述方法, 本文将定量分析结果转化为具体的教学资源, 帮助学生高效快速掌握重点词汇, 增强了语言学习的真实性和实用性。

4.2. 文本常见搭配分析

统计文本中的常见搭配及其频率如表 2 所示, 将文本常见搭配转化为具体的教学资源。

Table 2. Common collocations and their frequencies

表 2. 常用句型搭配及其频数

PATTERNS	COUNT
China witnessed increase	28
technique actors are makeup	23
pandas meet public	19
city witnessed growth	18
county tourism promoting industry	17
county guided enterprises upgrading	16
city leading development	16

续表

plum is role revitalization	15
improvement contributed	15
People celebrate festival	15
zone continued environment stability	15
summit gathered AI priorities participants	14
technology optimizes algorithms activities	14
it stressed role education yields systems	13
people celebrated Festival	12
number reached million	12
tourists enjoy scene Mountain	12
tourists visit spot	11
China made achievements	11
Chongqing built itself	10

1) 词汇教学：根据搭配频率统计结果，设计词汇搭配学习列表，并提供详细的解释、例句和相关练习，帮助学生在日常交流中灵活运用。创建包含高频搭配的词汇搭配表，并附上每个搭配的定义和例句。例如：China witnessed increase (例句：In recent years, China witnessed a significant increase in its GDP.); tourists visit spot (例句：Every year, millions of tourists visit this famous spot in Chongqing.).

2) 阅读理解练习：选择包含常见搭配的新闻段落，设计阅读理解问题，帮助学生在真实语境中理解和应用常见搭配。从新闻文本中选取包含高频搭配的段落，确保段落内容具有代表性和实用性。例如：China witnessed an increase in the number of tourists visiting Chongqing last year. The city’s growth in tourism has significantly contributed to its economy. 设计如“段落大意”、“搭配解释”和“内容细节”等问题，帮助学生理解段落内容。例如：What did China witness last year? How has tourism growth contributed to Chongqing’s economy?

3) 英语写作：根据常见搭配的主题设计写作任务，提升学生的写作能力。例如，可以让学生写一篇关于中国技术领域取得成就的文章，要求使用“China made achievements”、“technology optimizes”等搭配。设计与高频搭配相关的写作题目，例如：Discuss the achievements China has made in technology and how these advancements have optimized various activities.提供写作框架和搭配使用提示，帮助学生构思和组织文章。

4) 英语口语：利用常见搭配设计讨论话题，鼓励学生在课堂上进行讨论和交流。例如，可以让学生讨论重庆的发展和其在旅游业中的角色，使用“city witnessed growth”、“tourists visit spot”等搭配。设计与高频搭配相关的讨论话题，例如：Discuss the growth Chongqing has witnessed in recent years and how it has impacted the tourism industry. 组织小组讨论和辩论，鼓励学生使用高频搭配进行表达和交流。

借此本研究将量化分析结果转化为具体的教学资源，拓宽了英语专业培养应用型人才的路径。

4.3. 文本主题分析

本研究利用 LDA 主题模型进行分析，识别出五个主要主题：重庆旅游、重庆的企业发展、重庆国际贸易、重庆夜生活和西南交通，如下表 3 所示。

Table 3. News text topics**表 3.** 文本主题

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
tourism	enterprise	western	Chongqing	Chongqing
tourist	vehicle	centre	night	southwest
Chinese	trade	international	Chinese	railway
travel	growth	trade	food	speed
city	industry	investment	hotpot	construction
consumption	development	senior	southwest	kilometre
increase	energy	Chongqing	coffee	station
experience	economic	sixth	base	relief
accord	export	opening	product	bridge
million	billion	manufacturing	ingredient	high

1) 词汇教学：根据每个主题的高频词汇，设计重庆特色主题词表。创建包含各主题高频词汇的词汇表，并附上每个词的定义和例句。例如：主题 1：旅游：tourism (旅游)，例句：Tourism has become a major industry in Chongqing. 主题 2：企业：enterprise (企业)，例句：The enterprise has shown significant growth in recent years.

2) 阅读理解练习：选择包含这些高频主题词汇的新闻段落，设计阅读理解问题，帮助学生在真实语境中理解 and 应用这些词汇。从新闻文本中选取包含高频主题词汇的段落，确保段落内容具有代表性和实用性。例如：Chongqing has seen a significant increase in tourism, attracting millions of tourists each year. The city's vibrant night life and delicious hotpot are major attractions. 设计如“段落大意”、“词汇解释”和“内容细节”等问题，帮助学生理解段落内容。例如：What are the major attractions in Chongqing? How has tourism impacted the city's economy?

3) 英语写作：根据每个主题的高频词汇设计写作任务，提升学生的写作能力。例如，可以让学生写一篇关于重庆夜生活的文章，要求使用“Chongqing”、“night”、“food”、“hotpot”等词汇。设计与高频主题词汇相关的写作题目，例如：Discuss the impact of Chongqing's night life on its tourism industry. 提供写作框架和词汇使用提示，帮助学生构思和组织文章。

4) 英语口语：利用每个主题的高频词汇设计讨论话题，鼓励学生在课堂上进行讨论和交流。例如，可以让学生讨论西南地区交通的发展及其对经济的影响，使用“southwest”、“railway”、“construction”等词汇。设计与高频主题词汇相关的讨论话题，例如：Discuss how the development of the railway system in southwest China has impacted the region's economy. 组织小组讨论和辩论，鼓励学生使用高频主题词汇进行表达和交流。

通过上述方法，本文将 LDA 主题模型分析结果转化为具体的教学资源，为民办高校英语专业教学提供了科学依据和实用工具。

4.4. 语篇分析

本研究通过对语料库进行语篇分析，识别出文本的宏观结构、段落间的逻辑关系、主题的发展和信息组织方式。该类信息有助于开发具体的教学资源，如下表 4 所示。

Table 4. Discourse analysis statistical content, methods, and results
表 4. 语篇分析统计内容、方法及结果

统计内容	统计方法	统计结果
段落结构和逻辑关系	标注每个段落的主旨句、支持细节和逻辑关系	因果关系：33%
		对比关系：28%
		时间顺序：18%
		其他逻辑关系：21%
主题和子主题	识别主要主题和子主题，并分析其展开和发展	主题：重庆发展
		子主题 1：经济增长
		子主题 2：旅游业发展
		子主题 3：科技创新
信息组织方式	分析整体结构，包括引言、主体和结论	引言：8%
		主题：74%
		结论：18%

1) 词汇教学：根据语篇分析结果可知，逻辑关系为因果关系和对比关系的文本段落占一半以上，在词汇教学资源开发上，可以有针对性地拓展能够表达因果关系和对比关系的词汇和词组，例如，“Result in”，“Result from”，“Because of”等表示因果关系，以及“Compare with”，“Compare to”，“Relatively”等表示对比关系。形成一个专门的表达逻辑关系的词表，提高学生对逻辑关联词的使用能力。

2) 阅读理解练习：语篇分析中得到的具体主题和子主题，有助于学生理解文本的宏观结构和段落间的逻辑关系。选择一篇包含关键词“Chongqing’s economic growth”的长篇新闻文章，要求学生分析每个段落的主旨句和支持细节。练习识别段落之间的逻辑关系，如因果关系、对比关系等，帮助学生理解文章的整体结构。

3) 英语写作：更具体的子主题能够进一步明确英语写作练习的目标，培养学生撰写讲述中国故事文本的能力，同时，帮助学生组织和写作结构清晰、逻辑严密的文章。另外，分析语篇的信息组织方式，为学生示范逻辑更为严密，结构更为清晰的作文框架，引言、主题和结论的统计结果能够清晰地使学生把握文本各部分的词数占比。

4) 英语口语：语篇分析的结果可以作为训练学生在口语表达中使用结构框架，例如引言、主要观点、支持细节和结论的数据支撑。让学生在口语表达练习中使用逻辑连接词，使表达更有条理。设计关于“*How to develop Chongqing’s tourism better?*”的讨论，要求学生使用结构化的表达方式进行讨论和演讲。

本研究通过这些语篇分析教学资源开发，从词汇教学、阅读理解练习、英语写作和英语口语四个方面，帮助学生全面理解和掌握不同类型文本的整体结构特点，提高其综合语言应用能力。

5. 结语

本研究通过对《人民网英文版》(People’s Daily Online)收集的关于重庆的英语新闻文本进行定量分析，运用词频分析、常见搭配分析、LDA 主题分析和语篇分析三种方法，将英语新闻文本转化为教学资源，提高学生的英语应用能力。在文本常见搭配分析中，识别出文本中的常见搭配，开发相应的教学资源，增强学生的实际语言运用能力。LDA 主题模型分析提取出五个主要主题，基于主题设计的教学资源，使学生能够更好地理解和讨论与重庆相关的各类话题以及英语表达。总体而言，定量分析方法可以为民办

高校英语专业教学提供科学依据和实用工具, 能够显著提升教学效果和学生的语言能力, 实现学生语言综合能力的全面提升。具体的教学资源 and 案例不仅具有针对性和实用性, 还为英语教学提供了系统的指导, 增强了研究结果的应用价值。

参考文献

- [1] 李威. 语料库语言学在英语教育中的应用[J]. 中国教育学刊, 2020(S1): 81-82.
- [2] 杨子莹. 民办高校英语专业人才培养模式改革的研究[J]. 佳木斯职业学院学报, 2018(4): 414.
- [3] 任安莉. OBE 教育理念下民办高校英语专业发展路径探究[J]. 科教导刊, 2023(34): 17-19.
- [4] Smith, B. (2018) Addressing the Employability Crisis: Enhancing Language Skills through Corpus-Based Approaches. *International Journal of Language Studies*, **12**, 29-45.
- [5] Teubert, W. (2005) My Version of Corpus Linguistics. *International Journal of Corpus Linguistics*, **10**, 1-13.
<https://doi.org/10.1075/ijcl.10.1.01teu>
- [6] 田野, 李继民, 柳婷婷. 语料库驱动下的教师角色转变与定位[J]. 当代教育科学, 2012(7): 54-56.
- [7] Zhang, W. (2021) Using Corpora to Develop Teaching Resources in Private University English Programs. *Language Learning & Technology*, **25**, 1-20.
- [8] Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford University Press.
- [9] Bernardini, S. and Aston, G. (2004) Corpus-Aided Language Learning: A Flexible Model for Language Education. *ReCALL*, **16**, 375-390.
- [10] 李丽丽. 语料库在不同教学阶段的应用策略[J]. 语言教学与研究, 2020(3): 423-431.