

国内英语语料库研究综述—基于 CiteSpace 的可视化分析



王建*, 黄列梅

吉利学院外国语言与文化学院, 四川成都 641423

摘要: 随着计算机技术的发展, 语料库的建设与研究如火如荼地展开。在此背景下, 本研究基于 CNKI 中文期刊全文数据库, 借助 CiteSpace 文献计量可视化分析工具, 从发文量情况、作者合作网络、机构合作网络、关键词共现、聚类分析及关键词突变分析六个维度, 对 2002—2022 年间与英语语料库研究相关的 255 篇国内核心期刊论文进行了可视化分析。研究发现: (1) 这些论文主要发表在 13 种外语类核心期刊上, 且从 2016 年起发文量呈显著下降趋势; (2) 该领域的作者与机构合作网络较为稀薄; (3) 关键词分析表明, 研究者关注的话题多样, 不同时间段呈现出不同的研究热点, 第一个阶段 (2002—2007), 研究聚焦搭配及英语写作, 第二个阶段 (2008—2014), 研究热点为基于语料库的英语自主学习及网络资源的利用, 第三个阶段 (2016—2022), 研究热点转向词块、教学模式、学术英语及医学英语的探究。最后, 本文对未来的研究方向进行了展望。

关键词: 英语语料库; CiteSpace; 可视化分析; 研究热点; 研究趋势

DOI: 10.57237/j.jeit.2023.01.002

Focuses and Trends of Domestic Research on English Corpus: A Visual Analysis Based on CiteSpace

Wang Jian*, Huang Liemei

School of Foreign Languages and Culture, Geely University of China, Chengdu 641423, China

Abstract: Driven by the advances in computer technology, the construction of and research on corpus are in full swing. In this context, based on the CNKI database and aided by CiteSpace software, this study conducts a bibliometric visual analysis of 255 articles related to English corpus research and published in domestic core journals from 2002 to 2022. The analyses focus on six dimensions: publishing trend, authorial and institutional cooperation networks, keyword co-occurrence and cluster analyses, and analysis of the keywords with the strongest citation bursts. Results show that: (1) These articles are mainly published in 13 core journals on foreign languages, and the number of publications has shown a decreasing trend since 2016; (2) The authorial and institutional cooperation networks in this field are relatively thin; (3) Keyword analyses suggest that researchers have been dedicated to a variety of topics, and different research hotspots are presented in different time periods. The research focuses of the first stage (2002-2007) include collocation and English/second language writing. In the second stage (2008-2014), researchers zoom in on corpus-based autonomous English learning and the use of Internet learning

基金项目: 四川省哲学社会科学规划“四川省高校外国语言文学学科建设和发展项目”《大数据背景下汽车英语语料库建设与应用》(SC22WY008); 四川省经济发展战略研究会项目《大数据背景下汽车英语语料库建设与应用》(SCJFZL-YB-202230).

*通信作者: 王建, wangjian@bgu.edu.cn

收稿日期: 2023-01-15; 接受日期: 2023-02-16; 在线出版日期: 2023-03-02

<http://www.jeduit.com>

resources. The research hotspots of the third stage (2016-2022) turn to the exploration of lexical chunks, teaching modes, academic English and medical English. Finally, this article provides insights into the directions for future studies.

Keywords: English Corpus; CiteSpace; Visual Analysis; Research Focuses; Research Trends

1 引言

随着自然语言处理 (natural language processing) 等技术的出现及发展, 大型语料库的建设方兴未艾。语料库是用于学术研究和教学的语言数据的集合, 通常包含在计算机数据库中, 也通常被称为文本语料库。第一个系统组织的计算机语料库是布朗大学现代美式英语标准语料库 (The Brown University Standard Corpus of Present-Day American English, 简称为布朗语料库), 由语言学家亨利 库切拉 (Henry Kučera) 和 W 纳尔逊 弗朗西斯 (W. Nelson Francis) 于上个世纪 60 年代编制。国外比较著名的英语语料库有美国国家语料库 (The American National Corpus)、英国国家语料库 (British National Corpus)、当代美式英语语料库 (The Corpus of Contemporary American English) 和国际英语语料库 (The International Corpus of English) 等。国内英语语料库包括英语学习者语料库 (如广东外语外贸大学及上海交通大学联合建设的“中国学习者英语语料库”, Chinese Learners English Corpus, 简称 CLEC)、平行语料库 (如北京外国语大学建设的“汉英平行语料库”, English Chinese Parallel Corpora, 简称 PCCE) 以及特殊英语语料库 (如解放军外国语学院建设的“军事英语语料库”, Corpus of Military Texts, 简称 CMT)。Jan Svartvik 指出, 语料库具有很多优势[1]。首先, 语料库数据比基于内省的数据更客观, 可以很容易地被其他研究人员验证, 研究人员可以共享相同的数据, 而非编译自己的数据。其次, 语料库可用于研究方言、语域和风格之间的差异, 可提供特定语言项目出现的频率, 也可用于对语言特征进行全面分析, 不仅提供说明性示例, 而且还是一种理论资源。再次, 计算机化的语料库使世界各地的研究人员都可以访问数据, 也为非母语人士掌握目标语言提供支持。

目前, 语料库已被应用到许多领域。比如, 语料库衍生的词频列表是词典编纂者的基本工具。语料库也被运用到语言教学中, 使用索引作为语言学习工具是计算机辅助语言学习 (computer-assisted language learning) 的主要兴趣。近年来, 机器翻译也成为关注焦点, 翻译引擎运用双语及多语平行语料库并在自

然语言处理等技术的推动下比较准确地处理翻译任务。

由此可见, 语料库的建设有较大的实用意义, 及时揭示语料库相关研究的热点及趋势可以为该领域持续发展提供有益的参考。因此, 本研究将对国内英语语料库的研究进行综述。国内学者已经从不同视角对英语语料库的研究动态进行了梳理, 如基于语料库的英语教学研究[2-6]、医学英语语料库研究[7]以及语言研究方面的应用[8]。然而, 这些研究并不全面, 仅局限于某一领域, 无法揭示英语语料库研究的全貌。其次, 多数研究集中在 2015 年以前, 随着信息技术的进步, 语料库的应用亦获得了长足的发展, 是否有新的研究主题及热点出现不得而知。再次, 这些综述研究的研究方法、工具及呈现方式未能做到与时俱进, 目前尚无研究者利用可视化工具 (visualization tool) 对国内英语语料库相关研究进行分析。鉴于此, 本研究将弥补以上空白及不足, 旨在揭示国内英语语料库研究的热点及前沿, 并展望此领域未来的发展趋势。

2 研究设计

2.1 研究方法

本研究对采集到的数据进行可视化分析, 以知识图谱的形式呈现。知识图谱是应用数学、信息科学和计算机科学的跨学科领域, 是科学计量学领域的较新发展, 用于映射和分析科学知识, 从以前发表的科学研究文献中提取并直观地重组大量知识、发现新知识[9]。映射知识域 (mapping knowledge domains) 是一种以知识域为对象的图像, 并展示了科学知识的发展过程及其结构间的关系。它具有图谱和光谱的双重属性和特征——它不仅仅是一种视觉知识图, 也是连载的知识谱 (serialized knowledge pedigree), 显示了许多隐藏的、具有复杂关系的知识单元或知识组, 比如它们的网络、结构及交互等[10]。

2.2 研究工具

本研究的主要工具为美国德雷塞尔大学陈超美教授开发的 CiteSpace (引文空间)。研究人员可以运用

CiteSpace 探究某一学科的研究进展、研究热点和研究前沿[11]。目前, CiteSpace 已于 2022 年 12 月 22 日更新至 6.1.R6(64-bit)版本,本研究采用的是 6.1.R3(64-bit)基础版本,两个版本主要功能类似。CiteSpace 主要基于“共现聚类”思想,对科学文献的参考文献、关键词、主题词、学科、领域分类、作者、机构、国家、期刊等信息单元进行提取,并根据这些信息单元间的关联类型和强度进行重构,形成不同意义的网络结构(network,如关键词共现、作者合作、文献共被引等)。在网络结构中,节点(node)代表文献的信息单元,连线(link)代表节点间的联系或共现。CiteSpace 通过对节点、连线及网络结构进行可视化分析,找出某领域知识结构的隐含模式和规律。但对于中国知网(CNKI)来源的数据,CiteSpace 仅能完成作者、机构、国家及关键词的分析。

2.3 数据来源

本研究以“主题名=英语语料库”在中国学术期刊网络出版总库(CNKI)中进行高级检索,文章发表的时间跨度设定为 2002—2022 年,文献来源于 SCI、EI、北大核心、CSSCI、CSCD 及 AMI 检索的期刊。我们初步检索到 355 篇论文。为了数据的准确性,两位作者分别对检索到的文章进行筛选,筛除不符合文章发表时间段、文献来源及研究主题的论文,并删除重复出现的文章。此外,排除综述性及介绍性文章,只保

留实证研究论文。两位作者的判别一致率为 99.44%,存在分歧的地方经过协商解决。最终,我们确定符合标准的文章有 255 篇。虽然本研究收集的数据量相对较小,但仍符合 CiteSpace 的样本量标准[12]。

2.4 研究过程

首先,我们利用 Excel 2019 对筛选出的文献进行统计,包括 2002—2022 年间英语语料库相关研究每年的发文量以及每本期刊的总发文量,分析该领域的大体研究趋势和总体特征。其次,我们将 255 篇文章的 Refworks 文献格式在 CiteSpace 进行转换,因为软件不能直接分析 CNKI 来源的数据,可视化分析前必须进行数据转换。然后,我们对作者及作者机构的发文情况进行统计及可视化分析。最后,对文献中的关键词进行可视化分析,绘制出关键词共现(keyword co-occurrence)、关键词聚类(keyword cluster)、关键词突变(keyword citation burst)以及时间线(timeline)图谱,用以探索英语语料库领域中的研究热点和演进趋势。在可视化分析前,将 CiteSpace 的操作界面的节点类型(node types)选择为依次选择为“作者”、“机构”及“关键词”,时区选择为 2002—2022 年,时间切片(time slicing)为 1,切片最大阈值选择为 50,连线强度为 Cosine,路径不进行裁剪(pruning)。参数值设置好后进入可视化界面,我们根据需求调整节点最小阈值、节点形状、字体大小等,最终获得不同的图谱并进行解读。



图1 英语语料库相关研究发文量年度分布(2002—2022)

3 研究结果

3.1 发文量分析

从图1可以看出,2002—2022年时间段英语语料库

相关研究的年度发文量分布呈现较大波动,每年的平均发文量约为12篇。整体而言,从2002年至2012年,国内英语语料库研究呈快速增长趋势[13]。从2007年至2016年,年发文量超过平均值,10年间共有177篇文章出版,占文章总数的69.41%。这足以说明,在此时间段,基于英语语料库的研究受到了我国研究人员的极大

重视。发文量峰值出现在 2010 年 (N=28) 和 2012 年 (N=27)，其余年份发文量低于年发文量平均值。在 2016 年以后，年发文量降至平均值以下，且整体呈现下降趋势，表明英语语料库的研究热点逐渐退却。

本文作者还对各期刊发表英语语料库相关论文的情况进行了分析。统计发现，国内 84 种期刊发表了相关研究，21 年间每种期刊平均发文量约为 3 篇。表 1

罗列了发文量大于等于 5 篇的 15 种期刊，共计 153 篇，占有文章的 60%。如图 1 所示，除《现代教育技术》及《中国教育学刊》外，其余 13 种期刊均为外语类核心期刊，这些期刊共出版相关论文 143 篇 (56.08%)，占比较高的原因是本文的研究主题为“英语语料库研究”，与外语类核心期刊的匹配度最高。

表 1 国内 15 种核心期刊发文量分布

期刊名	N	年份	期刊名	N	年份
外语电化教学	30	2004	现代外语	8	2003
外语界	14	2009	山东外语教学	8	2008
解放军外国语学院学报	14	2003	中国外语	7	2009
外语教学与研究	13	2002	外语与外语教学	6	2005
外语教学	12	2002	外语研究	5	2007
外语教学理论与实践	10	2010	现代教育技术	5	2009
外语学刊	8	2002	中国教育学刊	5	2016
西安外国语大学学报	8	2014			

注：本文所有表格中 N 代表文章或关键词数量，年份指初始发表年份。

值得一提的是，上海外国语大学主办的《外语电化教学》期刊总共发表了 30 篇相关论文，占有论文数量的 11.76%。我们对该期刊的发文情况进行了详细分析，结果发现，从 2004 年开始到 2012 年，该期刊每年的刊文量为 1 到 4 篇，2013-2015 三年没有刊发相关论文，但从 2016 年到 2020 年，每年的刊文量又恢复至 1 到 4 篇。30 篇论文的话题覆盖面较广，涉及通用英语教学方面的语料库研究（如英语词汇教学、写作教学、修辞教学及课堂话语等方面的语料库研究）[14-16]、基于语料库的特殊用途英语教学探究（如数学英语）[17]以及利用语料库对公示语英译规范化的研究等[18]。

3.2 作者合作网络分析

借助 CiteSpace 中显示的有关作者的统计分析发现，255 篇文章共涉及 196 位作者，其中 15 位作者发文量大于等于 3（见表 2），25 位作者发文量为 2 篇，这 40 位作者的发文数量占文章总数的三成以上，其余作者发文 1 篇，这说明国内关于英语数据库的研究呈现出单核心的发展模式，少数核心成员占据重要地位，发挥着举足轻重的作用。发文量最大的作者为王立非教授，5 篇论文中 1 篇是关于商务英语词汇名化的语料库[19]，另外 4 篇均基于语料库探究大学英语写作的语言特征[20-23]。

表 2 高发文量作者信息

作者名	N	初始发表年份	作者名	N	初始发表年份
王立非	5	2006	戈玲玲	3	2011
许家金	4	2007	徐昉	3	2008
何安平	4	2007	张岩	3	2008
陈建生	3	2004	娄宝翠	3	2013
王芳	3	2008	刘芹	3	2010
汪兴富	3	2008	刘永兵	3	2010
杨永林	3	2004	Mark	3	2008
李文中	3	2003			

CiteSpace 的作者合作网络分析共形成了 127 个节点，47 个连接，合作密度 (density) 为 0.0057。由图 2 可以看出，作者合作网络较为稀薄，主要形成了王立非—张岩、张济华—高钦—王蓓蕾、李端阳—王志军、

高文成—梅德明以及杨永林—李鸣 5 组研究合作网络，其中王立非、张岩两位研究者合作较为突出，共合作发文 3 篇，其余 4 组合作发文两篇。但总体来看，研究者之间的合作网络没有明显形成。

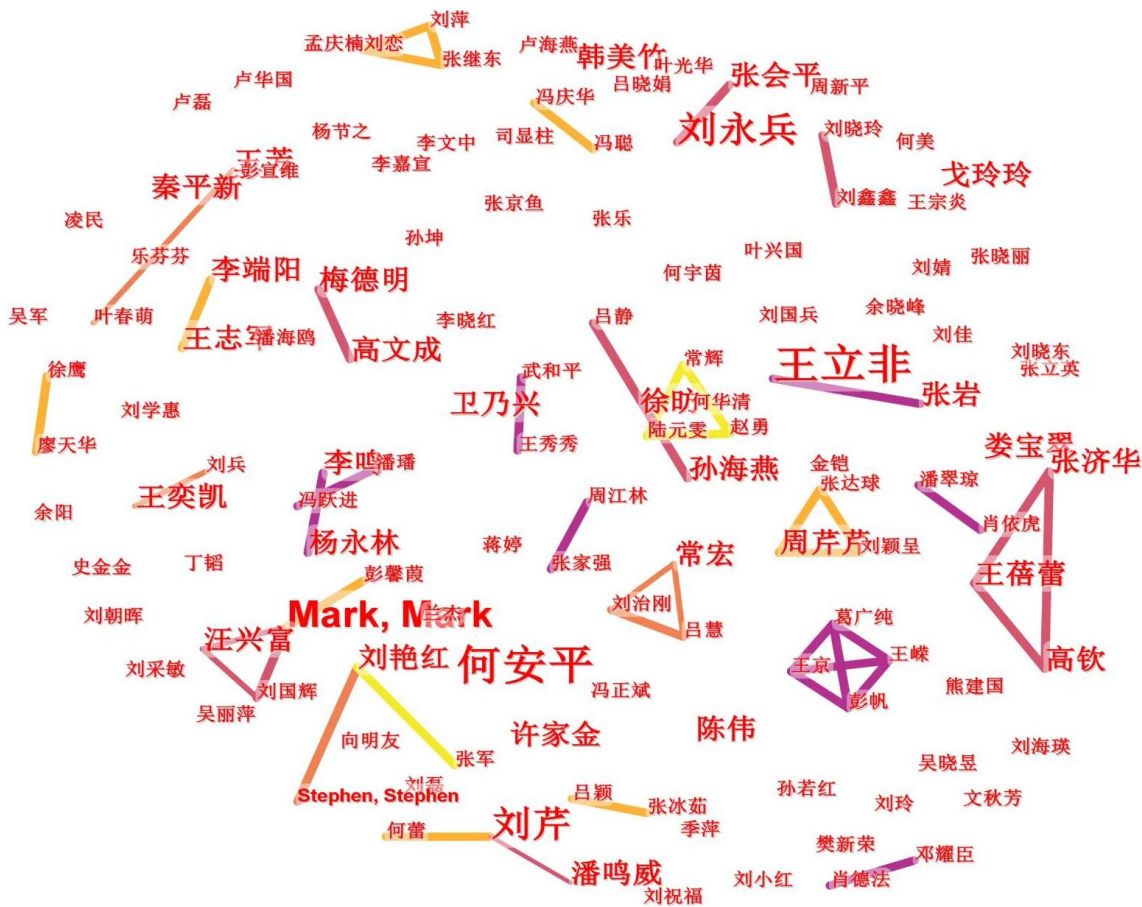


图 2 作者合作网络图谱

3.3 机构合作网络分析

统计发现，所有研究人员来自 90 个机构，表 3 列出了出现频次为 3 次及以上的机构。从表格中可以看出，上海外国语大学出现频次最高，为 12 次，说明该校学者对英语语料库相关研究做出的贡献较大。

表 3 主要机构及发文量信息

机构名	N	机构名	N
上海外国语大学	12	上海理工大学外语学院	4
河南师范大学外国语学院	8	上海对外经贸大学	3
上海交通大学外国语学院	6	北京航空航天大学	3
中国外语教育研究中心	5	东北师范大学外国语学院	3
对外经济贸易大学英语学院	4	南京大学外国语学院	3
北京外国语大学	4	解放军外国语学院	3
上海交通大学	4	华南师范大学	3

由图 3 可知，机构合作网络图谱呈现出明显的聚集效应，说明英语语料库研究领域成果集中在少数研究机构。节点之间连线的粗细反映了研究机构之间合作的强弱。从文献归属来看，绝大多数成果来源于知名外国语

大学 and 高校外国语学院，较少有科技公司或科研机构对英语语料库展开研究。此外，各大高校与其他企业、科研机构或单位的合作较少，这一现象反映出近二十年英语语料库领域的学术研究来源较为单一。



图4 英语语料库研究文献关键词共现图谱

[24]。CiteSpace 软件可以分析出同一领域文献中每个关键词出现的频次及其中心性 (centrality)，并绘制出反映关键词频次高低及中心度的谱图，因此能够直观

地呈现该领域的研究热点[25]。将 CiteSpace 的操作界面的节点类型选择为“关键词”，最终得到了英语语料库研究文献的关键词共现图谱（见图 4）。图谱中出现字号最大的关键词为“语料库”，说明该关键词在 255 篇论文中出现的频率最高。图中还出现了其他字号相对较小的关键词，如词块、语法化等，说明学者对这些相关话题有关关注。

表 4 列出了出现频次大于等于 3 的 18 个关键词及

其在 2002—2022 年间首次出现的年份。可以看出，语料库出现频次最高，为 136 次，这与关键词共现图谱的结果相得益彰。从表中的关键词，我们可以得出结论：学者们对基于英语语料库的英语教学研究（如英语教学、英语写作、自主学习等）、语言特征研究（如词汇特征、搭配、语法化等）以及特殊用途英语（如商务英语、中医英语、医学英语等）产生了浓厚的兴趣。

表 4 高频关键词信息

关键词	N	年份	关键词	N	年份	关键词	N	年份
语料库	136	2002	词块	6	2010	中医英语	3	2003
英语教学	11	2002	应用	5	2008	教学	3	2005
搭配	9	2002	口语教学	4	2007	二语习得	3	2010
学术英语	8	2006	写作教学	4	2008	学习者	3	2010
英语写作	7	2004	标注	4	2008	词汇特征	3	2010
词汇	7	2005	大学英语	4	2009	语法化	3	2010
英语	6	2003	中介语	4	2010	中国英语	3	2012
自主学习	6	2007	英语口语	4	2010	医学英语	3	2012
商务英语	6	2008	词语搭配	4	2011	美国英语	3	2012

CiteSpace 还可以计算出每个关键词的出现历史（history of appearance，即某关键词每年出现的频次）以及临近节点（neighbouring node，即与某关键词联系最紧密的关键词）。以“语料库”为例，该关键词在 21 年间均有出现（见图 5），出现频次的峰值出现在 2011 年（N=17），说明其为当年的研究热点。此外，软件计算出该关键词的临近节点有 203 个，如作用、对比研究、搭配、僵化等。

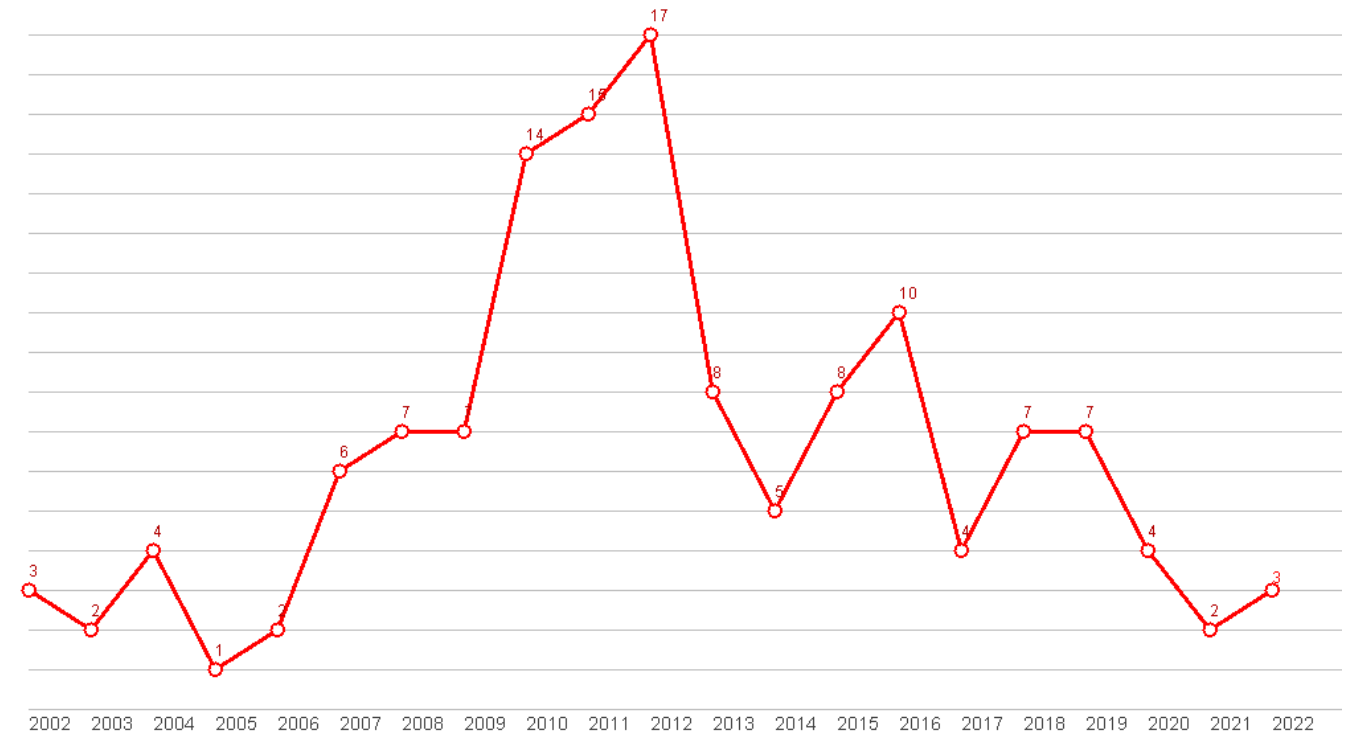


图 5 关键词“语料库”出现历史

3.5 关键词聚类分析

研究热点是特定学术领域学者关注的重点，它反映了特定时间内研究者在实地讨论的主要问题。如前文所述，作为学术论文的重要组成部分，关键词通常用于研究某一领域的研究热点，因为它们体现了论文的精髓。基于此，本文利用 CiteSpace 进行了关键词聚类分析，该分析可将图 4 中联系较为紧密的关键词进行汇总形成聚类，更加直观反映英语语料库的研究热点。我们采用 CiteSpace 软件的对数似然率 (Log-Likelihood Ratio, 简称 LLR) 聚类算法代替余弦定理计算相似度。LLR 越大的关键词在某一聚类中越具有代表性。聚类分析结果显示，整个网络一共包括 33 个聚类。在导出结果时，我们将聚类数量设置为 10，即图 6 中显示的 Cluster 0#到 Cluster 9#，图谱种色块表示聚类区域。CiteSpace 依据网络结构和聚类的清晰度，提供了模块值 (Modularity Q, 简称 Q 值) 和平均加权轮廓值 (Mean Weighted Silhouette, 简称 S 值) 两项指标，当 Q 值>0.3 时，聚类结构就是显著的[26]，当 S 值达到 0.7 就可以认为聚类是令人信服的[27]。本研究结果显示，Q 值为 0.6876，S 值为 0.9679，满足了基本要求，10 聚类标签均有效。



图 6 英语语料库研究文献关键词聚类图谱

图 6 中各聚类按数值从小到大所包含的关键词量依次递减。本文将 10 个聚类的相关数据汇入表 5，包含聚类类别 (cluster ID)、关键词数量、S 值及聚类内代表性关键词，每个聚类提供最核心的 3 个关键词及其对数似然率。

表 5 关键词聚类汇总信息

聚类	N	S 值	聚类内代表性关键词
0#	77	0.985	语料库 (11.5)、英语教材、搭配 (5.33)
1#	35	0.965	搭配 (23.05)、类联接、词语搭配 (9.05)
2#	27	0.928	学术英语、结构特征 (10.13)、英语标题 (5.03)
3#	24	0.985	中国英语 (10.98)、对应语料库、句法分析 (5.45)
4#	21	0.941	应用 (17.92)、英语 (11.83)、分析 (5.86)
5#	17	0.954	词汇 (16.67)、类符、形符 (5.86)
6#	17	0.952	英语教学、口语教学 (10.98)、数学英语 (5.45)
7#	15	0.971	商务英语 (19.12)、商务合同、高频词 (6.23)
8#	11	0.975	错误分析、英语写作 (13.22)、错误诊断 (6.53)

我们也利用 CiteSpace 制作了时间线图谱。该图谱可直观呈现各聚类之间的相互联系以及每个聚类内部各节点的历时演变进程[28]。如图 7 所示，每个聚类的标签均为共现网络中的关键词，这些关键词按照其在对应时间内所出现的年份，在所属的聚类中铺展开来，显示每个聚类里关键词的发展情况。具体而言，各个聚类根据大小自上而下排列，每个聚类的主要关键词按照其出现的时间先后顺序自左向右排列，连线表明各关键词间的传承关系。实线两端分别表示聚类内部关键词出现及结束的时间。

从图谱中可以看出，最突出的关键词为“语料库”，该关键词在时间线上的节点最大，连线最密集，从 2002 年到 2022 年均呈现出强劲的发展态势与传承关系，说明语料库研究仍然没有失去热度。比如，在 2022 年，语料库与课程思政思辨要素相结合的研究就证明了语料库研究的创新发展。相比之下，其他关键词的发展态势跟传承性较弱，在时间线上的节点较小，连线较少。比如关键词“被动语态”在时间线上的节点很小，颜色很淡，其连线仅有几条，说明英语语料库领域研究人员对该话题的研究兴趣不浓厚。

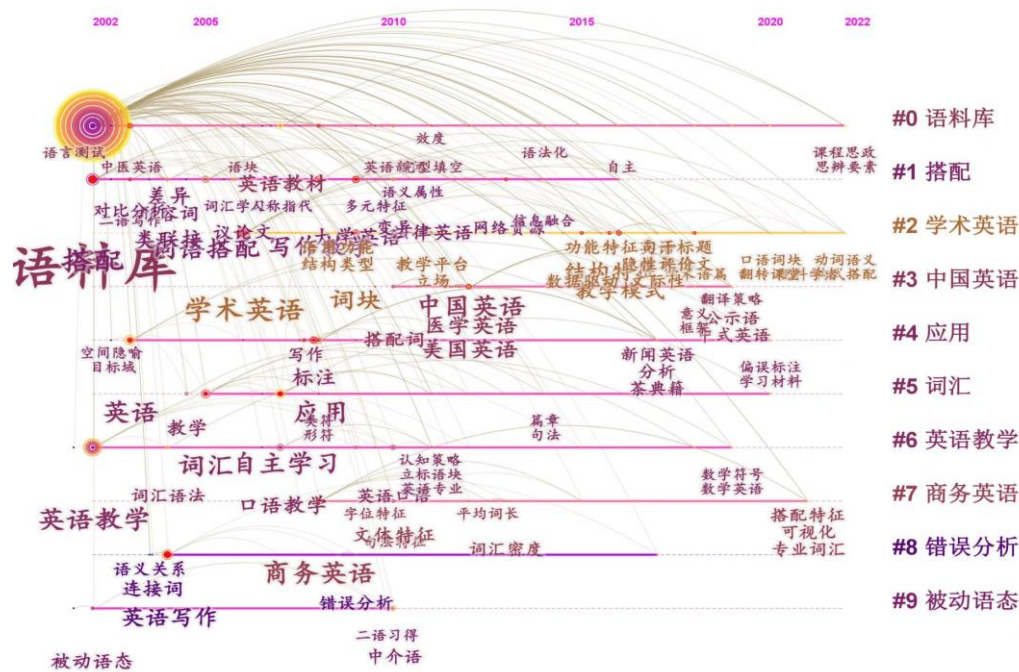


图 7 英语语料库研究文献关键词聚类时间线图

3.6 关键词突变分析

关键词突变指在较短的时间内某个关键词突然出现或使用频次突然增多的现象[29]。CiteSpace 通过突变检测算法（burst detection）从论文的核心信息（如标题、摘要、关键词等）中提取出突变术语，然后绘制关键词突变图谱，用于探究某研究领域的热点话题及研究前沿[25]。

Top 15 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2002 - 2022
搭配	2002	2.68	2002	2007	<div></div>
英语写作	2004	2.03	2004	2010	<div></div>
自主学习	2007	1.89	2007	2011	<div></div>
英语	2003	1.53	2003	2006	<div></div>
词块	2009	1.5	2016	2022	<div></div>
教学模式	2016	1.39	2016	2020	<div></div>
词汇	2005	1.39	2005	2008	<div></div>
标注	2008	1.3	2008	2012	<div></div>
学术英语	2006	1.26	2016	2022	<div></div>
二语写作	2003	1.2	2003	2006	<div></div>
写作	2008	1.2	2008	2009	<div></div>
词汇教学	2008	1.2	2008	2009	<div></div>
网络资源	2013	1.16	2013	2014	<div></div>
中医英语	2003	1.15	2017	2018	<div></div>
医学英语	2012	1.15	2018	2019	<div></div>

图 8 英语语料库研究文献关键词突变图谱

CiteSpace 绘制的关键词突变图谱中列出了英语语料库前 15 个突变关键词、关键词在 2002—2022 年间首次出现的年份、突变强度 (strength)、突变的开始和结束时间。红色线条表示关键词突变的持续时间段 (见图 8)。如图所示, 21 年间除 2015 年以外, 其他年份均出现了突变关键词, 说明总体上英语语料库研究热点持续存在。突变强度值 (2.68) 最高的是关键词“搭配”, “英语写作”(2004—2010)、“词块”(2016—2022) 和“学术英语”(2016—2022) 的突变持续时间最长。在 2002—2007 年间, 搭配、英语/二语写作等是英语语料库研究的热点, 这一阶段研究人员主要利用语料库对学习者的语言进行语言学分析。在 2008—2014 年间, 研究者们更关注基于语料库的英语自主学习及网络资源的利用, 该阶段基于语料库的英语学习引起了学者们的广泛关注。而在 2016—2022 年间, 研究热点转向词块、教学模式、学术英语以及医学英语的探究, 语料库对英语教学的辅助作用成为新的研究方向。

4 结论及展望

本文主要借助 CiteSpace 软件对 2002—2022 年间发表在国内核心期刊上的、与英语语料库相关的 255 篇论文进行了可视化分析。从发文情况来看, 相关文章主要发表在国内 13 种外语类核心期刊上, 且自 2016 年以来, 该领域的论文数量呈现下降趋势。在合作网络方面, 我们发现该领域研究者及机构合作网络较为稀薄, 尚未形成明显的合作网络。从关键词分析来看, 研究者关注的话题多样, 形成了大量的关键词, 关键词聚类与突变分析也表明了该领域研究的热点及发展趋势。

基于可视化分析结果, 我们认为, 要实现英语语料库研究领域的可持续发展, 未来研究可从以下几个方面着手。首先, 255 篇文章的作者主要是高等学校英语教师, 为了解决合作网络稀薄的问题, 老师们可展开跨学科、跨机构合作, 比如可以与科技公司的技术人员携手利用语料库解决现实问题。英语教师不必单枪匹马, 应展开大量的跨学科研究, 这也符合学界倡导的跨学科教学研究的要求[30]。与技术人员的跨界合作可以更加深入、更加科学地进行相关研究。其次, 研究者们需要探寻新的研究话题, 不局限于基于语料库的英语教学研究, 需要更新研究视角。比如, 在特殊用途英语的相关研究中, 有研究者对基于语料库的

医学英语和学术英语进行了探索, 但尚未针对其他专门用途英语 (如汽车英语、金融英语、体育英语等) 建设语料库或进行基于语料库的分析。通过不断探寻新的研究方向, 研究者们将推动英语语料库研究迈向新的发展阶段。

总之, 本研究利用可视化分析工具仅对近 20 年国内英语语料库研究进行了初步分析, 研究结果并不全面, 还存在一些不足之处, 但能够大体上揭示该领域的研究热点及方向。我们希望, 本研究能够为致力于英语语料库研究的学者们提供具有一定价值的参考。

参考文献

- [1] Svartvik, J. Directions in Corpus Linguistics [Z]. Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991, Berlin. New York: De Gruyter Mouton, 1992. DOI: 10.1515/9783110867275
- [2] 周卫京, 王苏婉. 我国基于语料库的英语口语教学研究 20 年综述 (2002~2021) [J]. 当代外语研究, 2022, (4): 123-132.
- [3] 洪敏芳, 王昕玥, 李平. 语料库在英语教学中的应用研究综述 [J]. 语言与文化论坛, 2022, (2): 265-274.
- [4] 方秀才. 基于语料库的英语教学与研究综述: 成就与不足——根据 22 种语言学类 CSSCI 来源期刊近 30 年的统计分析 [J]. 外语电化教学, 2012, 145 (3): 19-24.
- [5] 陈玉芬. 国内基于语料库的英语教学研究综述——基于 2011-2013 年中国知网核心期刊的文献调查 [J]. 北方工业大学学报, 2014, 26 (4): 60-65.
- [6] 安雪花. 语料库在外语写作教学中的应用研究评述及展望 [J]. 西安外国语大学学报, 2013, 21 (2): 54-58.
- [7] 康勇、王翠莲. 医学英语语料库研究现状综述 [J]. 辽宁医学院学报 (社会科学版), 2015, 13 (4): 136-138.
- [8] 穆晓莉, 尹转云. 语料库在语言研究方面的应用综述 [J]. 西安外国语学院学报, 2001, (3): 21-26.
- [9] Chen, C. *Mapping Scientific Frontiers* [M]. London, England: Springer-Verlag, 2003.
- [10] Chen, C. *How to Use CiteSpace* [M]. Victoria, British Columbia, Canada: Leanpub, 2015.
- [11] 马蓉, 刘玉飞. 我国外语教学研究的热点、前沿及演进——基于中国知网学术期刊文献的共词可视化分析 [J]. 北京第二外国语学院学报, 2021, 43 (4): 19-32+115.

- [12] Chen, C. Science mapping: A systematic review of the literature [J]. *Journal of Data and Information Science*, 2017, 2 (2), 1-40. DOI: 10.1515/jdis-2017-0006.
- [13] 刘皓. 语料库在中国的研究发展与趋势 [J]. 现代交际, 2017, 454 (8): 170-172+169.
- [14] 黄若妤. 语料库指引的词汇教学 CIVI 模式建构 [J]. 外语电化教学, 2008, (2): 52-57.
- [15] 杨劲松, 罗永胜. 语料驱动的微观修辞操作与英语写作教学研究 [J]. 外语电化教学, 2009, (6): 18-24.
- [16] 刘学惠. 英语师范生课堂话语的建库、分析和应用 [J]. 外语电化教学, 2006, (5): 41-45.
- [17] 金娜娜, 彭宣维. 数学英语的主要特点与数学符号表征——兼谈数学话语语料库研制的基本内容 [J]. 外语电化教学, 2019, (3): 28-35.
- [18] 周芹芹, 张达球, 刘颖呈. 基于语料库的校园公示语英译规范化研究 [J]. 外语电化教学, 2019, (2): 67-74.
- [19] 王立非. 商务英语词汇名化的语料库考察及批评分析 [J]. 外语电化教学, 2012, 145 (3): 3-9.
- [20] 王立非, 张岩. 大学生英语议论文中疑问句式使用的特征——一项基于中外学习者语料库的对比研究 [J]. 解放军外国语学院学报, 2006a, (1): 43-47.
- [21] 王立非, 张岩. 基于语料库的大学生英语议论文中的语块使用模式研究 [J]. 外语电化教学, 2006b, (40): 36-41.
- [22] 王立非, 张岩. 大学生英语议论文中高频动词使用的语料库研究 [J]. 外语教学与研究, 2007, (2): 110-116, 160-161.
- [23] 王立非, 陈功. 大学生英语写作中分裂句的特征——基于语料库的考察 [J]. 外语教学与研究, 2008, (5): 362-367+401.
- [24] Su, X., Li, X., & Kang, Y. A bibliometric analysis of research on intangible cultural heritage using CiteSpace [J]. *Sage Open*, 2019, 9 (2), 1-18. DOI: 10.1177/2158244019840119.
- [25] 李杰, 陈超美. CiteSpace: 科技文本挖掘及可视化 [M]. 北京: 首都经济贸易大学出版社, 2017.
- [26] Newman, M. E. Fast algorithm for detecting community structure in networks [J]. *Physical Review E*, 2004, 69(6), 066133. DOI: 10.1103/PhysRevE.69.066133
- [27] Rousseeuw, P. J., & Kaufman, L. Finding groups in data [J]. *Series in Probability & Mathematical Statistics*, 1990, 34(1), 111-112. DOI:10.1002/9780470316801.
- [28] 陈悦, 陈超美, 刘则渊, 胡志刚, 王贤文. CiteSpace 知识图谱的方法论功能 [J]. 科学学研究, 2015, (2): 242-253.
- [29] 李红满. 国际翻译学研究热点与前沿的可视化分析 [J]. 中国翻译, 2014, (2): 21-26, 127.
- [30] 谭春辉, 刁斐, 李瑶, 周一夫. 科研人员个体感知影响跨学科科研合作意愿机理 [J/OL]. 情报理论与实践. <http://kns.cnki.net/kcms/detail/11.1762.G3.20221122.1136.004.html>