

基于CiteSpace的国内英语搭配实证研究可视化分析

王雪纯

四川大学外国语学院, 四川 成都

收稿日期: 2023年6月28日; 录用日期: 2023年8月4日; 发布日期: 2023年8月15日

摘要

搭配研究一直是国内英语学习研究领域关注的一个焦点, 也是长期影响二语学习的一个重要因素。文章基于CiteSpace这一可视化软件, 对国内英语搭配实证研究的进展进行了较为系统且全面的回顾, 探寻了该领域的相关热点和难点, 并展望了未来的研究趋势, 进而在有限的思考内为今后的研究提供一定的思路。研究发现, 国内英语搭配研究的热点领域包括英语教学、语料库研究和心理学分析; 在该领域研究发展的过程中, 存在研究主题单一化、历时性研究稀缺、跨学科性不足这几个特征。未来研究要注重丰富研究主题, 充分利用学科交叉融合, 并促进研究方法的本土化。

关键词

英语搭配, 实证研究, CiteSpace, 计量分析

A Visualized Analysis of Domestic Empirical Research on English Collocation Based on CiteSpace

Xuechun Wang

College of Foreign Languages and Cultures, Sichuan University, Chengdu Sichuan

Received: Jun. 28th, 2023; accepted: Aug. 4th, 2023; published: Aug. 15th, 2023

Abstract

Collocation research has always been a focus in the field of domestic English learning, and it is also an important factor affecting second language learning for a long time. Based on CiteSpace, a visual software, this paper aims to systematically and comprehensively review the progress of domes-

tic empirical research on English collocation, explore the relevant hot and difficult points in this field, look forward to the future research trend, and then provide some ideas for future research within limited thinking. It is found that the hot areas of domestic research on English collocation include English teaching, corpus research and psychological analysis; in the process of research development in this field, there are several characteristics: single research theme, lack of diachronic research and lack of interdisciplinary. Future studies should focus on enriching research topics, making full use of interdisciplinary integration, and promoting the localization of research methods.

Keywords

English Collocation, Empirical Research, CiteSpace, Econometric Analysis

Copyright © 2023 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

在语义层面,搭配(collocation)是指两个或更多的词习惯连用或共现的现象[1]。以往的大量研究显示,搭配是二语习得研究的一大热点。学习者的搭配行为和搭配能力是语言能力的重要组成部分,很多国内外语言学者也都跻身于这个领域的研究中来。

词汇的搭配并不是随意的,而是存在某种必要的联系。从这个角度来说,词汇搭配的相关研究应将目光放在词汇之间的这种关联上来。目前,在我国的搭配研究中,对搭配的定义未达成一个广泛的共识,其研究使用的术语也存在明显的差异,包括“搭配”、“习语”、“程式语”、“语块”等,这给搭配研究的系统分析带来了一些困难。随着搭配研究的深入,其研究对象和研究方法也呈现出多样化的趋势,特别是相关的实证研究。在我国,英语搭配的实证研究一直是搭配研究的焦点,有关研究成果层出不穷,为英语教学的改进提供了理论依据和现实参考。鉴于此,本文旨在通过系统性的回顾和梳理,以可视化分析软件 CiteSpace 为基本工具,对中文社会科学引文索引(CSSCI)所有有关英语搭配的期刊论文进行基本的回顾分析,探讨总结研究相关进展,主要回答了以下几个问题:1)国内英语搭配实证研究的总体情况如何?2)国内该领域的研究特征是什么?3)国内该领域研究的发展趋势是什么?

2. 研究工具与数据采集

2.1. 研究工具

本研究使用 CiteSpace 为基本工具进行计量分析,其版本为 CiteSpace 5.8.R3 (64 bit),是由美国德雷赛尔大学(Drexel University)陈超美等人研制开发的知识可视化工具。作为目前最为流行的知识可视化工具之一,CiteSpace 已经成为文献可视化分析研究的常客,其性能和使用也得到持续的优化,不断为各类知识图谱应用研究提供坚实的技术保障。自 2005 年陈超美团队率先在中国命名和引入科学知识图谱(mapping knowledge domains)以来,有关该领域的科学计量学研究便在我国扎根下来,并取得了长足的发展。科学知识图谱是以知识域(knowledge domain)为对象,显示科学知识的发展进程与结构关系的一种图像[2]。CiteSpace 这一软件应用 Java 语言开发,可以对特定学科或领域的文献进行集中考察,根据其绘制的知识图谱,研究者可以清楚看到某一研究领域的节点信息,进而分析该领域发展的潜在动力机制和

趋势, 能够帮助研究者更好地理解某一学科领域的研究情况, 抓住研究重要主题等关键信息, 为该研究领域的总结和归纳提供高效的方法。正如刘则渊总结的那样: “一图谱春秋, 一览无余; 一图胜万言, 一目了然” [2] (p. 244)。

基于以上信息, 本研究将对国内英语搭配实证研究的相关文献进行计量分析, 从作者、机构、研究主题等方面绘制图谱, 并对其研究的热点难点以及发展趋势进行详细说明。

2.2. 数据采集

本次研究的数据来源为中文社会科学引文索引(CSSCI), 它是由南京大学中国社会科学研究评价中心开发研制的数据库, 用来检索中文社会科学领域的论文收录和文献被引用情况, 具有较高的学术质量和代表性, 可以有效保证本研究的科学性和可信性。由于搭配、语块、程式语、习语等术语存在诸多交叉、重叠, 学界对其定义和划分尚无定论[3], 且在数据检索时, 输入“英语搭配”进行所有字段检索得出的数据过少(仅 12 条), 在考虑设置条件等多重原因之后, 为提高数据检索过程中的准确性和科学性, 本文以“搭配”、“语块”、“程式语”、“习语”这四个关键词和“英语”两两组合为检索词进行所有字段检索, 得出 167 条数据。在对这些文献进行逐一审查后, 排除综述等非实证性研究 81 篇, 最终得到英语搭配实证研究文献 86 篇。这一数据的总量是比较少的, 但是考虑到 CSSCI 来源期刊存在发文机构固化的现象, 即由于 CSSCI 来源期刊以“名校名家”、“地缘关系”、“学缘关系”作为论文发表的参考因素而导致的发文头部机构固化现象[4], 本研究数据较少的现象就得到了解释。值得注意的是, 这里的实证研究文献既包括定量研究, 也包括如个案分析等定性研究, 同时, 在将文献数据导入 CiteSpace 软件处理之前, 为了减少研究误差, 再次对筛选后的文献进行了二次甄别, 在最大程度上保证了研究的准确性和真实性。

3. 国内英语搭配实证研究的时间分布与研究力量

3.1. 文献发表的时间分布

通过对 86 篇文献的归纳整理, 得出文献的时间分布情况(如图 1 所示), 其中包括语言学文献 83 篇, 教育学文献 3 篇。由图 1 可知, 国内英语搭配实证研究的文献数量总体上来说是呈现缓慢增长趋势的, 虽然文献发表数量时增时减, 但不难看出, 实证研究的文献数量在慢慢增加。在这个过程中, 出现以下情况: 第一, 国内英语搭配的实证研究文献数量增长时有起伏, 每年相关的文献产出不稳定的, 文献发表数量在 2011 年到达顶峰(12 篇); 第二, 从近十年来看, 即 2011~2021 年, 每年的文献产出数量呈现下降趋势, 这说明该领域研究的后劲不足, 需要提高科研发展水平, 激发该领域的研究动力。从整体来看, 我国有关英语搭配的实证研究还是比较少的, 特别是高质量高水平的学术文献还处于一个稀缺的状态, 这需要学者们加大研究力度, 增强研究后劲, 提升该领域的科研发展水平。

3.2. 文献发表的研究力量

在对该领域研究的作者和机构进行计量分析后, 呈现出以下的情况。

以“作者”为检索项在 CiteSpace 中进行分析得到以下结果。如图 2 所示, 图中的节点表示各个作者发文的情况, 节点和文字的大小代表的是各自的期刊发文数量, 节点和文字越大, 该作者的期刊发文数量越多。各个节点之间的连线代表了他们的合作关系, 连线的粗细表示的是合作关系的深度, 单个作者没有节点代表他们彼此之间不存在合作关系。由图 2 可知, 在期刊发表文献较多的学者有张萍、蔡金亭、钟志英等, 而单个作者的节点代表他们发文数量少, 且与其他作者之间不存在合作关系。从图中可以看出, 各个学者之间的联系不是特别紧密, 有相当多的学者选择单独研究, 且有合作关系的作者之间的合

作程度也并不深入，呈现出多分散、少合作的作者关系。

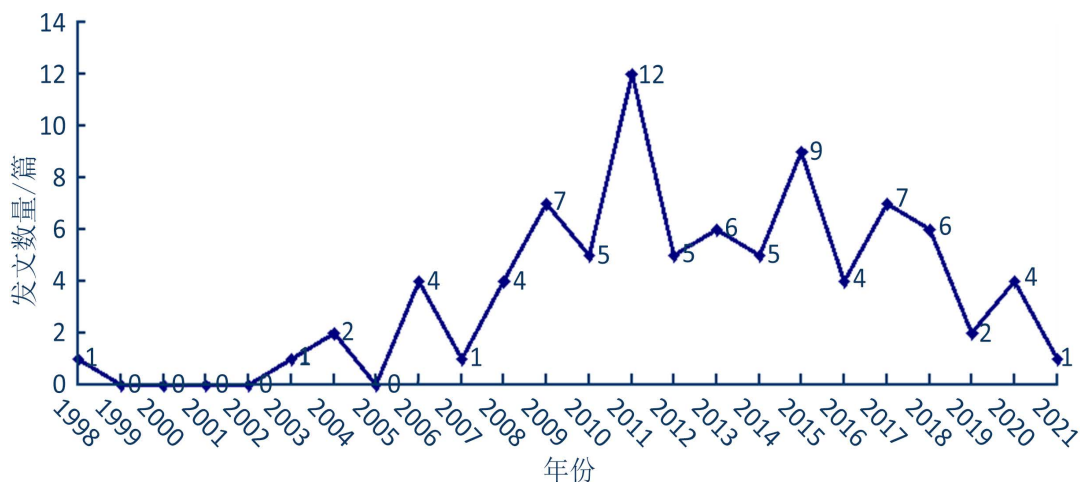


Figure 1. Temporal distribution of domestic empirical research literature on English collocation from CSSCI
图 1. CSSCI 来源国内英语搭配实证研究文献的时间分布

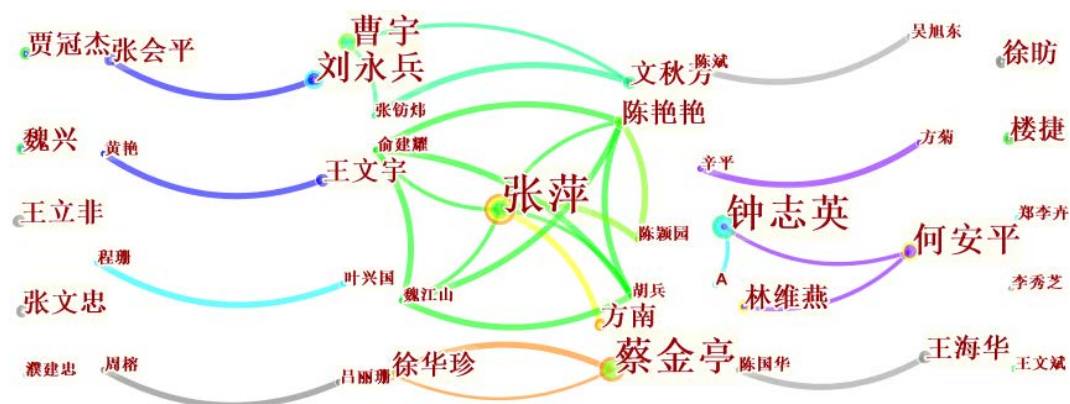


Figure 2. Distribution of author relationships in domestic empirical research on English collocation
图 2. 国内英语搭配实证研究的作者关系分布

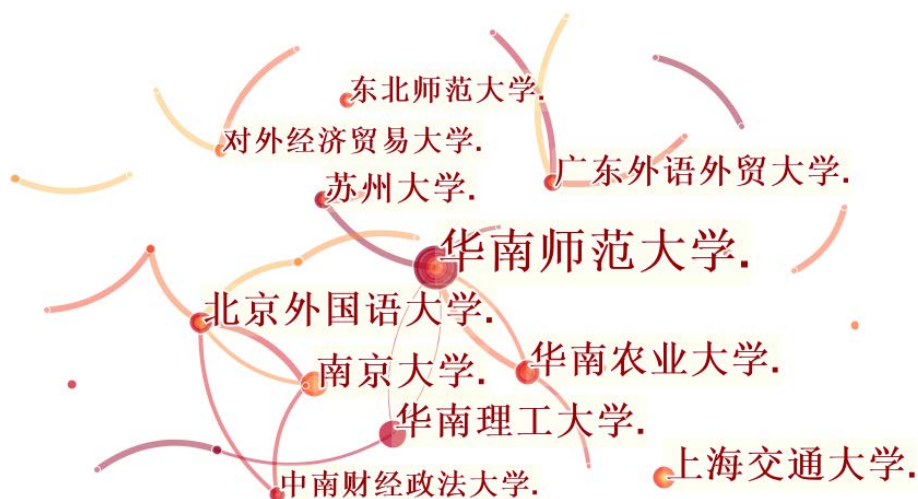


Figure 3. Distribution of institutional relationships in domestic empirical research on English collocation
图 3. 国内英语搭配实证研究的机构关系分布

在考察发表文献机构的分布时,将“机构”设置为检索项并得到以下情况(如图3所示)。由图3可知,在所有发文机构中,有几所非常具有代表性且在该研究领域表现突出,它们分别是华南师范大学、北京外国语大学、南京大学、华南农业大学、华南理工大学和上海交通大学。其中,华南师范大学表现尤为突出,还有一些其他的机构表现也很不俗,如广东外语外贸大学、东北师范大学、苏州大学等。但值得注意的是,机构之间的合作关系不是特别密切,且有些机构完全处于游离状态,如上海交通大学,并没有和其他机构进行深入合作。

4. 国内英语搭配实证研究的进展回顾

4.1. 研究主题

在仔细阅读了本研究所涉及的所有文献的标题、中英文摘要、关键词和正文之后,本文对国内英语搭配实证研究的主题进行了简要分类,包括了英语教学、语料库、心理学分析三个方面,分类情况如下表1所示。需要说明的是,在确定研究的主题类别时,本文参考了一些文献的分类方法[3][5]。由于很多文献会出现类别交叉的现象,即一篇文献可归属于两类甚至三类,所以在分类时本文采用了以下方法:比较文献的各类特征,挑选最突出最鲜明的特征作为文献归类的依据。比如,如果一篇文章是涉及语料库使用的英语学习研究,那么它将被归入语料库的类别之中。

Table 1. Thematic composition proportion of domestic empirical research on English collocation

表 1. 国内英语搭配实证研究的主题构成占比

语料库	心理学分析	英语教学
44%	12%	44%

由表1可知,在国内英语搭配的实证研究中,语料库和英语教学所占的比重相当,是国内研究的主要方向,也是最为常见的研究主题。孙海燕[6]通过对比中国英语学习者和本族语者在形容词搭配方面的不同模式,从语义角度探析学习者在形容词搭配使用中的典型特征。其研究基于两种类型的语料库,即中国学习者英语语料库 CLEC (Chinese Learner English Corpus)和本族语语料库 LOB (The Lancaster-Oslo/Bergen Corpus)、JDEST (上海交通大学科技英语语料库)、COBUILD (柯林斯伯明翰大学国际语料库)。杨素香[7]选取 BNC (British National Corpus)笔语部分 1,482,815 词次的学术医学英语和 3,678,647 词次的传记英语为研究语料,探讨语料库证据显示的不同语类如医学英语和传记英语的语义序列。而有关心理学分析的研究占比较小,和其他两个主题相比,心理学分析稍显弱势。在有关搭配的心理分析研究中,多数研究采取了以下思路:基于产出实验,对学习者的搭配行为的心理表征进行了具体分析。李更春和贾冠杰[8]对中国英语学习者进行了两项实验,即自定步速逐词阅读和逐句阅读实验,研究将学习者分为中等水平和高等水平两组受试,采用 2×2 混合实验得出程式语结构的完整性没有影响两组被试对程式语的心理表征模式,并总结出了中国英语学习者对程式语的心理加工与表征模型。

从 1998 年到 2021 年,国内英语搭配的教学实证研究主要包括中国英语学习者的搭配加工[9][10]和产出[6][11][12];语料库研究涵盖语义分析[7][13][14]、语境分析[15][16]以及迁移分析[17][18];心理学分析主题主要涉及认知过程[19][20]。

4.2. 研究方法

基于对文献的细读,本文将国内英语搭配实证研究的研究方法分为以下四类:语料库、问卷、访谈和任务型测试,得到的结果如下表2所示。由表2可知,国内英语搭配实证研究的研究方法主要包括四种,其中基于语料库的研究占比最大,其次为任务型测试,这两种研究方法共同构成了搭配研究的主流。

除此之外, 问卷和访谈也是常见的研究方法, 但是占比不高。在基于语料库的研究中, 大部分是依靠已知的权威语料库, 如 BNC 和 CLEC。贾光茂[21]利用 CLEC 语料库和翻译测试两种研究方法, 从定量和定性两种视角分析了 31 名南京某高校英语专业三年级学生的动名搭配产出, 并进行了产出错误的概念迁移分析, 试图回答“英汉概念化方式差异是否会引起中国英语学习者产出错误的动名搭配”这一问题。其研究结果显示, 汉英识解方式差异是导致动名搭配错误的重要原因。还有一些研究因没有现成的语料库, 就采用了自建语料库的方法。例如, 蔡金亭和徐华珍[11]运用历时方式收集了 32 名大学生的口语语料, 他每隔 4 个月进行一次口语测试, 共进行 5 次, 系统分析了动名搭配产出错误的显著性变化。

值得注意的是, 有些研究会将多种方法结合, 如将问卷和访谈相结合, 进而达到多角度多方位的研究效果, 在对它们进行分类的时候, 为了使得分类达到预期的效果, 本文仅考虑其研究方法中的最基本的方法。

Table 2. Research method composition proportion of domestic empirical research on English collocation

表 2. 国内英语搭配实证研究的研究方法构成占比

任务型测试	语料库	访谈	问卷
32%	44%	8%	16%

4.3. 关键词共现

关键词是一篇文章的核心, 代表着文章的精华和要点。通过对关键词的考察, 能够发现至今为止研究的热点和难点, 如果某个关键词反复出现, 则代表其为该领域研究的重点和趋势所在。本文使用 CiteSpace 的“关键词”检索项进行操作, 绘制出国内英语搭配实证研究的关键词共现图谱, 如图 4 所示。由图 4 可知, 该领域研究反复出现的关键词有语块、语料库、搭配、程式语、动名搭配、语义韵、类联接等。为了进一步了解英语搭配实证研究领域的更多信息, 本文利用 CiteSpace 的时间线算法, 根据文献的标题、摘要、关键词、正文、引用等信息绘制了时间线图, 如图 5 所示, 这能帮助研究者更直观、更清晰地看到该领域发展的变化和趋势。

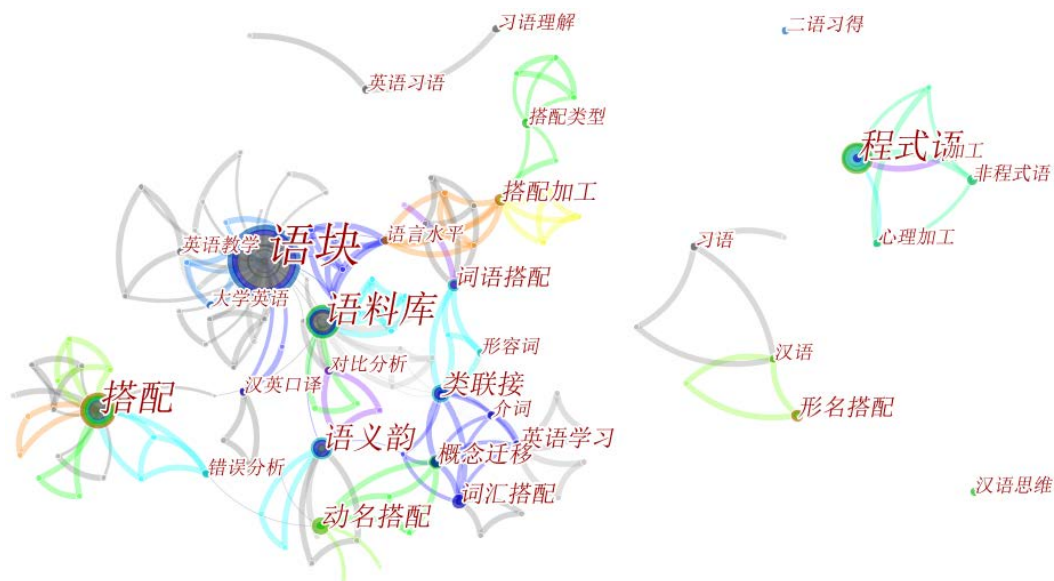


Figure 4. Keyword co-occurrence knowledge map of domestic research on English collocation

图 4. 国内英语搭配研究的关键词共现知识图谱

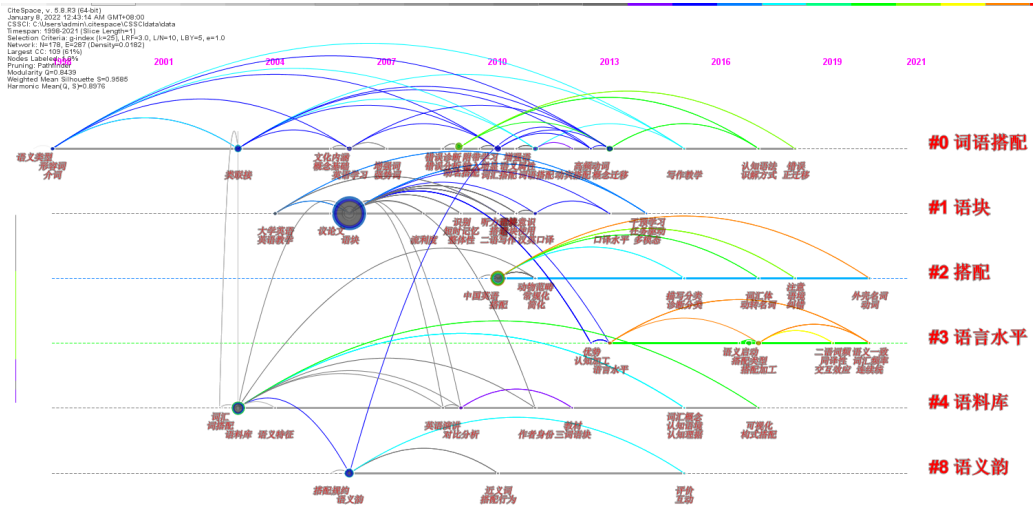


Figure 5. Keyword timeline map of domestic empirical research on English collocation
图 5. 国内英语搭配实证研究的关键词时间线图

图 5 的时间线图的右边是关键词的聚类，它们的序号代表其重要程度，聚类越重要其序号越靠前。这些聚类标签包含的关键词信息按照首次出现的时间顺序依次铺开，最终呈现在时间线的坐标轴上。关键词之间的连线代表的是它们之间的关系，这些连线的深浅粗细就是它们之间联系的紧密程度。由图 4、图 5 可知，词语搭配聚类是该领域最初的研究重点，之后有关搭配的术语不断变化，“语块”、“习语”、“程式语”等术语被应用到国内搭配研究中；搭配关注的主题也由最初的英语教学不断细化，涵盖听力、口译、写作、认知、语义、语境等各个方面，丰富了该领域的研究成果；随着大数据和人工智能技术的深入发展，语料库研究得到越来越多的关注和重视，从上世纪 90 年代开始，国内语料库研究开始兴盛起来，这也为搭配的语料库研究提供了有力的理论支持[22]。

5. 国内英语搭配实证研究的特征

5.1. 研究主题单一化

在国内所有研究英语搭配的实证研究中，大多数都是围绕英语教学中的搭配问题进行的实验，且其中绝大部分研究的研究对象都是中国大学生，即英语学习的高水平者，目前针对这个群体的研究还有很多，不仅是在搭配领域，在二语习得的其他领域也屡见不鲜。但是，很少有研究关注到中低英语水平者的搭配产出情况。研究对象的单一进一步加剧了主题的统一性，而在这之中，有关动名搭配的研究占比最大[21] [23] [24] [25]，其他的词语搭配研究较少，形容词、介词搭配等研究寥寥无几。除此之外，有关习语和程式语的研究也比较少，其中有很大一部分研究关注的是学生对习语的理解[26] [27] [28]。

5.2. 历时性研究稀缺

在考察所有文献的研究思路和方法之后，本文发现，几乎所有研究都是采用横向的共时性研究思路，很少有研究会考虑到历时研究的方法。缺乏历时性的考察，研究者就不能很好地关注到搭配的动态情况，会使得该领域的研究变得单薄，减弱了研究的创新性和思辨性。针对这一点，有少部分研究采取了共时和历时研究相结合的研究方法，这不失为未来研究的一个较好的研究思路。除此之外，各研究使用的方法步骤也较为相似，语料库和任务型测试的研究方法虽然有效，但不一定适合所有的研究。在选取研究的方法时，研究者应该考虑到方法的适配性、有效性和可行性。

5.3. 跨学科性不足

就目前的情况来看,国内英语搭配的实证研究主要向着心理学的领域发散,研究中国英语学习者在进行搭配加工时的认知过程[8][29],梳理学习者在这个过程中所表现出来的心理表征或者思维特征。这些是对搭配研究的丰富和拓展,是跨学科研究的突出体现。但是就现有文献来看,搭配领域的跨学科步伐处于停滞状态,并没有进行进一步的发散,其呈现的心理学相关的研究也比较单一,研究缺乏后劲,在文献发表上表现出总量小、种类少的特点。

6. 结论与展望

回顾我国英语搭配实证研究的发展历程,本文从时间分布、研究力量、研究方法、热点难点、研究特征等方面入手,对国内英语搭配的实证研究领域的学术论文总体情况进行文献计量和可视化分析,简要地梳理了该领域研究的重要成果。考虑到本次研究最终所得数据比较少,可能会对后面结果分析造成影响,使得数据结果的显著性较弱,所以本次研究只是一次初步的尝试,希望会对后面的研究提供可参考的思路。

在归纳总结了该领域的研究大致情况后,本文提出了以下几点展望和思考。

第一,丰富研究主题。未来研究者们可以尝试跳开英语教学的单一主题,往更加具有创新性和价值力的研究主题发展,使得该领域的研究更加多元和深入,从而为该领域的研究注入新的生机。比如,研究的范围可以拓宽到地域的对比研究等等。研究的对象也不仅仅是限定在大学生身上,可以是高水平学术研究者、企业、政府等。这样的选择可以打破目前该领域研究的固化现象,重新激发研究的活力,进而丰富该领域的研究成果。

第二,充分利用学科交叉融合。打破传统学科之间的壁垒,促进基础学科、应用学科交叉融合,促进文理渗透等多形式交叉。英语搭配的研究不应该只限定在语言学的研究范围之内,要将语言的研究和其他领域的研究结合起来,这一点几乎适用于所有的语言研究。除了和心理学的交叉,英语搭配的研究还可以和计算机、医学、商业贸易等相结合,打开搭配研究的新视野,适应新时代中的研究趋势和风向。

第三,促进研究方法的本土化。关于英语搭配的研究,很多研究方法和框架都是来自于国外的研究成果,在设计研究方案时,要注意根据研究者自己的实际情况,基于本土研究的特征和要求,科学评判研究方法和框架的有效性和可行性,积极合理地做出调整,提升研究的信度和效度。

参考文献

- [1] 源可乐. 论英语教学型词典中的搭配信息[J]. 现代外语(季刊), 2001, 24(3): 259-266.
- [2] 陈悦, 陈超美, 刘则渊, 等. CiteSpace 知识图谱的方法论功能[J]. 科学学研究, 2015, 33(2): 242-253.
- [3] 张群, 杨连瑞, 陈颖, 陈士法. 二语搭配研究进展[J]. 现代外语(双月刊), 2021, 44(1): 123-132.
- [4] 甘琳, 李刚. CSSCI 来源期刊发文机构固化现象的成因与影响——以图情档期刊为例[J]. 理论研究, 2021, 41(2): 56-66.
- [5] 王璐璐, 戴炜栋. 二语习得研究方法综述[J]. 外语界, 2014(5): 29-37.
- [6] 孙海燕. 基于语料库的学生英语形容词搭配语义特征探究[J]. 现代外语(季刊), 2004, 27(4): 410-418.
- [7] 杨素香. 语料库驱动的语义序列研究——以搭配框架为例[J]. 外语教学理论与实践, 2012(4): 28-35.
- [8] 李更春, 贾冠杰. 英语程式语心理表征模式研究[J]. 西藏大学学报, 2013, 28(2): 174-180.
- [9] 张晓鹏, 张晓瑾. 频率和方向偏好对英语动宾搭配加工的影响研究[J]. 外语与外语教学, 2020(2): 16-26.
- [10] 方南, 张萍. 同译性与二语词频对英语搭配加工的影响[J]. 现代外语, 2019, 42(1): 85-97.
- [11] 蔡金亭, 徐华珍. 中国大学生英语形名搭配使用中的母语迁移历时研究[J]. 现代外语, 2018, 41(6): 817-828.
- [12] 秦悦. 对大学生英语词语搭配能力的调查和思考[J]. 上海大学学报(社会科学版), 2004, 11(5): 79-84.

- [13] 秦新平. 法律英语增强语的语义属性及词语搭配调查——一项基于法律汉英平行语料库的研究[J]. 新疆大学学报(哲学·人文社会科学版), 2011, 39(2): 153-156.
- [14] 卫乃兴. 基于语料库学生英语中的语义韵对比研究[J]. 外语学刊, 2006(5): 50-54.
- [15] 段士平. 复杂系统理论框架下中国大学生英语口语模糊限制语块使用研究[J]. 外语界, 2015(6): 45-53.
- [16] 程珊, 叶兴国. 不同认知语境下词汇概念搭配建构的认知理据研究——基于商务英语和普通英语语料库的“PREJUDICE”个案研究[J]. 中国外语, 2015, 12(4): 69-79.
- [17] 张会平, 刘永兵. 英语词汇学习与概念迁移——以常用动词搭配、类联接与语义韵为例[J]. 外语与外语教学, 2013(6): 38-43.
- [18] 魏兴, 张文霞. 汉语思维对英语搭配联系构建和搭配词汇输出的影响[J]. 现代外语(双月刊), 2017, 40(2): 232-243.
- [19] 桑紫林, 张少林. 中国英语学习者语块认知加工优势研究[J]. 外语教学理论与实践, 2013(2): 36-41+88.
- [20] 钟志英, Teresa Cadierno. 中国和丹麦 EFL 学习者对英语高频程式语的认知加工研究[J]. 外语教学理论与实践, 2015(2): 17-23.
- [21] 贾光茂. 中国英语学习者动名搭配错误的概念迁移分析[J]. 外语教学理论与实践, 2017(1): 30-35+66.
- [22] 黄水清, 王东波. 国内语料库研究综述[J]. 信息资源管理学报, 2021, 11(3): 4-17.
- [23] 张文忠, 杨士超. 中国学习者英语语料库中动名搭配错误研究[J]. 解放军外国语学院学报, 2009, 32(2): 39-44.
- [24] 刘国兵, 张孝莲. 语料库驱动视角下学术英语动词搭配配价研究[J]. 外语电化教学, 2021(1): 105-111.
- [25] 王海华, 周秀娟. 中国英语学习者动名词搭配行为的发展特点研究——语料库驱动的研究方法[J]. 外语学刊, 2009(6): 59-62.
- [26] 吴旭东, 陈斌, 黄丽辉. 中国学生对英语习语的理解: 习语类型与二语水平的作用[J]. 外语教学与研究(外国语文双月刊), 2006, 38(3): 196-201.
- [27] 张雷, 俞理明. 心理类型在中国学生英语习语理解中的作用[J]. 现代外语(季刊), 2011, 34(2): 171-177.
- [28] 唐玲. 中国大学生英语习语理解策略研究[J]. 外语教学理论与实践, 2009(2): 10-16.
- [29] 徐昉. 英语专业学生二语限时写作提取语块的思维特征[J]. 外语与外语教学, 2010(1): 22-26.