Modern Linguistics 现代语言学, 2022, 10(5), 842-848
Published Online May 2022 in Hans. http://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2022.105111

Hans汉斯

基于CiteSpace的国内语料库翻译学研究的可视化分析

——以CNKI核心期刊(1993~2021)为例

肖　丽

上海海事大学外国语学院，上海

收稿日期：2022年4月12日；录用日期：2022年5月9日；发布日期：2022年5月19日

摘　要

以近30年(1993~2021) CNKI数据库检索的在核心期刊上发表的有关语料库翻译学文献为数据来源，借助文献计量软件CiteSpace对语料库翻译学研究从3个方面进行系统分析，在此基础上绘制语料库翻译学研究知识图谱，结果显示近30年的语料库翻译学发文量呈波动趋势，具体可划分为三个阶段，同时提出了一些问题与展望。

关键词

语料库翻译学，文献计量工具CiteSpace，可视化分析

Visual Analysis of Translation Studies in Domestic Corpus Based on CiteSpace

—Taking CNKI Core Journals (1993~2021) as an Example

Li Xiao

School of Foreign Languages, Shanghai Maritime University, Shanghai

Received: Apr. 12th, 2022; accepted: May 9th, 2022; published: May 19th, 2022

Abstract

Taking the relevant corpus translation studies published in the core journals retrieved in the CNKI database in the past 30 years (1993~2021) as the data source, and using the bibliometric software

文章引用: 肖丽. 基于 CiteSpace 的国内语料库翻译学研究的可视化分析[J]. 现代语言学, 2022, 10(5): 842-848.
DOI: 10.12677/ml.2022.105111

CiteSpace to systematically analyze the corpus translation studies from three aspects, the knowledge map of corpus translation studies is drawn on this basis. The results show that the volume of corpus translation studies has fluctuated in the past 30 years, which can be divided into three stages, and some questions and prospects are raised at the same time.

Keywords

Corpus Translation Studies, Bibliometric Tool CiteSpace, Visual Analysis

1. 引言

语料库应用于翻译研究始于 20 世纪 80 年代末期，主要是作为一种工具，进行语言对比和翻译批评方面的研究。90 年代初期，平行对齐的语料库数据开始应用于机器翻译[1]。1993 年，英国学者 Mona Baker 发表了“语料库语言学与翻译研究：启示和应用”一文，将语料库研究方法引入翻译研究，此文标志着语料库翻译研究的开端[2]。随后，国内专家学者也开始致力于语料库翻译学的研究，语料库翻译学逐渐成为国内翻译学和语料库语言学的研究热点，引起了专家学者的广泛关注[3]。如今，国内的研究已经走过了近 30 个春秋，也有了长足的发展。

本研究主要是基于语料库翻译学的定义，研究对象是中国知网核心期刊上 1993~2021 年的语料库翻译学相关文献，通过定性与定量相结合的研究方法，从而探索语料库翻译学的研究规律和趋向。

2. 数据分析

CiteSpace (陈超美，2015)是通过知识图谱的方式探究某一领域发展的最新趋势[4]。选取中国知网数据库学术期刊数据子库，设置主题词“语料库翻译”，锁定文献来源为 SCI 来源期刊、北大核心、CSSCI，经过检索和人工筛查，剔除了书评、征稿启事等非学术期刊文章，共得到 1063 篇文章[5]。最早可见郑宝山于 1993 年在《上海科技翻译》发表的“美国《化学文摘》翻译及机器翻译”，因此本研究数据的时间设置为 1993~2021 年。本研究将所保存的相关论文题录信息为原始数据(data)、导入 CiteSpace 进行数据转化、存于新建项目并进行数值设置：时间跨度为 1993~2021，时间切片(time slicing)设置为一年，节点类型选择关键词(Keyword)，提取每个时区中关键词出现频率前 50 的数据构建图谱。

2.1. 发文量

为了研究语料库翻译学发文量的时间分布，将在中国知网检索的文章按年进行统计分析生成国内语料库翻译学研究文献数量走势图(图 1)。如图 1 所示，我国语料库翻译学研究的发文量从 1993 年至 2021 年整体呈波浪式增长。根据增长速度的不同，语料库翻译学在 1993~2021 年的发展可以划分为萌芽阶段(1993~2001)、发展阶段(2002~2008)、成熟阶段(2009~2021)。萌芽阶段发文量基数极小，增速缓慢。在增长速度最快的发展阶段，其发文量在低缓期发展的基础上增幅迅速加大。在随后的成熟期中，尽管增长速度有所下降，却达到了量的突破，发文量在 2017 年达到了巅峰。这一时期是语料库翻译学繁荣发展阶段，且持续时间将远长于前两个阶段[6]。

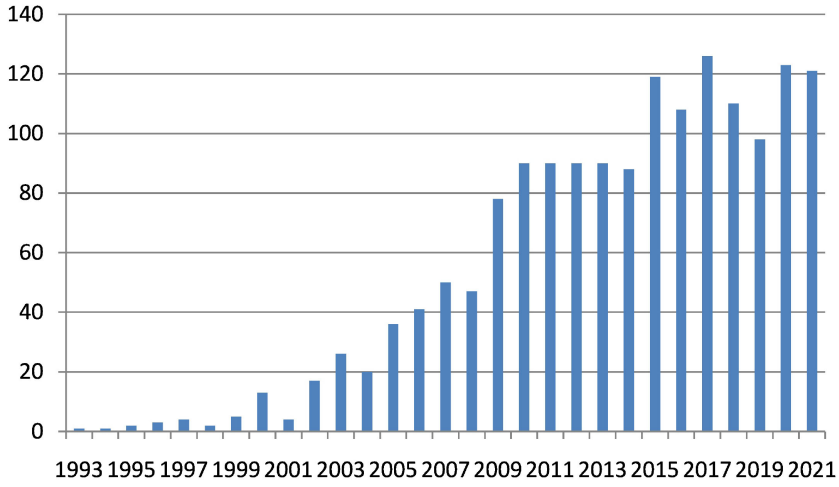

Figure 1. Annual publication volume of corpus translation studies
图 1. 语料库翻译学的年度发文量

2.2. 研究者

研究者对学科的发展演进意义重大，在推动学科发展过程中不可或缺。从图 2 可以看出我国语料库翻译学研究领域核心作者共线知识网络密度较低，表明作者之间的合作程度较低，仅形成了几个较有影响的合作团队，如王克非团队、胡开宝团队、刘泽权团队等。其中以王克非团队规模最为壮大，其团队成员主要有胡显耀、黄立波和秦洪武等。这一领域高产作者以单核心发展模式发展，未出现双核心或多核心模式，而单核心模式的团队合作过分依赖于核心人物，网络牢固性较差，不利于学科的发展和进步，甚至会出现研究停滞的情况。

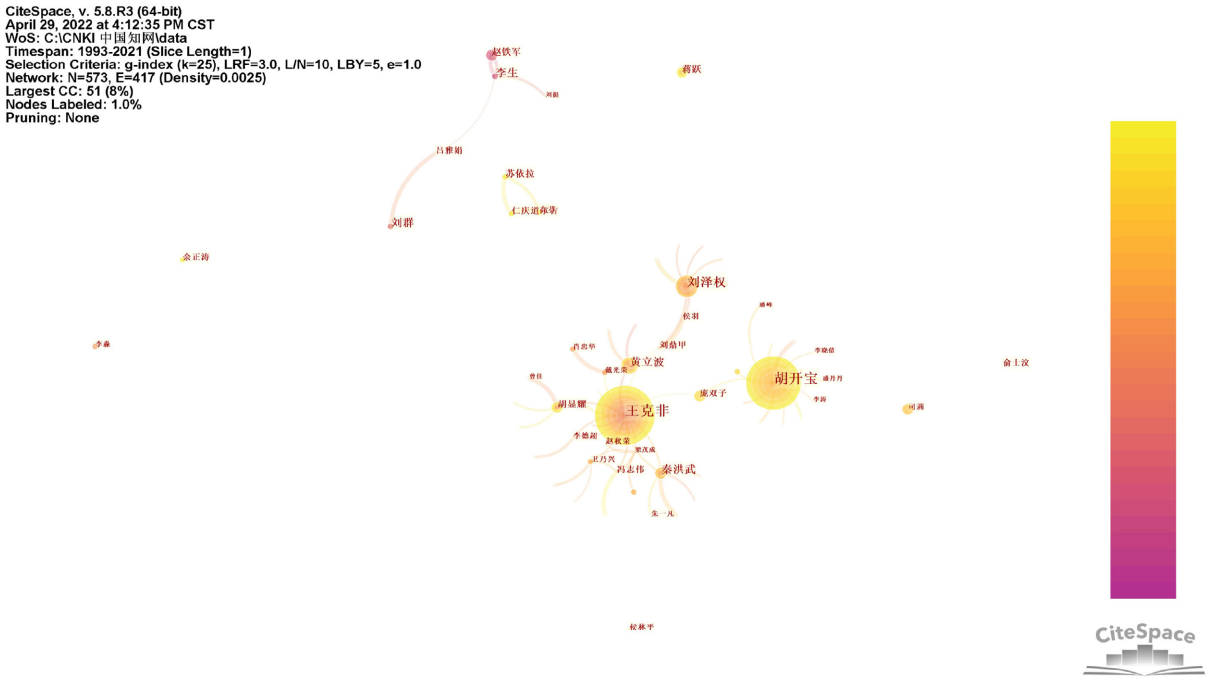

Figure 2. Author collaboration map
图 2. 作者合作图谱

为进一步了解核心作者的发文情况，将发文量排名前 10 位高产作者整理出来，如表 1 所示。由表 1 可知，王克非从 2004 年就致力于语料库翻译学的研究，至今以发表相关论文 36 篇，无论从发文量还是发文起始年份来说，王克非均在该领域独树一帜，是国内该领域的领军人物，为语料库翻译学的发展做出了重大贡献。此后，胡开宝、刘泽权等纷纷进军语料库翻译学领域，这一研究队伍逐渐壮大。

Table 1. Information of 10 highly-productive authors
表 1. 10 位高产作者信息

序号	作者	发文量	发文起始年份
1	王克非	36	2004
2	胡开宝	31	2009
3	刘泽权	19	2008
4	刘群	16	1999
5	李生	14	1996
6	秦洪武	13	2004
7	赵铁军	12	1996
8	黄立波	12	2008
9	苏依拉	10	2017
10	戈玲玲	10	2011

2.3. 研究领域

关键词共线分析是文献计量学常用方法之一，根据各个关键词之间的联系反映某个领域内部关系，探索其发展动态与趋势。图 3 中十字图形的大小即为关键词的频次，频次越高，十字越大。连线越粗则代表关键词之间的紧密程度[7]。CiteSpace 依据网络结构和聚类的清晰度，提供了模块值(Q 值，即 Modularity Q)和平均轮廓值(S 值，即 Mean Silhouette)两个指标，当 Q 值 > 0.3 时，聚类结构就是显著的；当 S 值达到 0.7 就可认为聚类是令人信服的。图 3 左上角的数据显示 Q 值 = 0.9406，S 值 = 0.9779，因此该聚类图谱的聚类结构十分显著，且结果令人信服[8]。

基于以上描述从图 3 中可以发现，在国内语料库翻译学研究中，节点度值最高的关键词是“语料库”，由此可见语料库之于语料库翻译学研究的意义。

将图 3 前 8 个聚类的相关数据汇入表 2。表 2 中“聚类内代表性关键词”一栏中的第一个关键词为该聚类名称，因为其代表性最强。同时提取了每一聚类的前五个代表性最强的关键词，从而帮助定位国内语料库翻译的研究领域。

通过对表 2 中的关键词进行整合分析，分析发现，国内学者的研究主要聚焦在三个方面：

第一个方面：译学研究语料库建设研究。语料的搜集是任何基于语料库研究的基础，对研究进展的顺利与否有着决定性的作用[9]。

第二个方面：翻译语言特征研究。在译学语料库的协助下，研究者探讨了翻译语言特征这一问题，如对翻译共性的研究，包括显化、隐化和范化，除此之外还有对语义韵的研究以及其他语言特征的研究等[10]。这有利于研究者发现翻译的规律，对翻译学习有一定的指导作用。

第三个方面：译者风格研究。不同的译者具有不同的翻译风格，对译者风格的研究不仅能加深对原文的理解，而且通过对两个或两个以上译者的译文进行对比分析，对某译者常用的翻译策略及动因的探究也大有裨益。

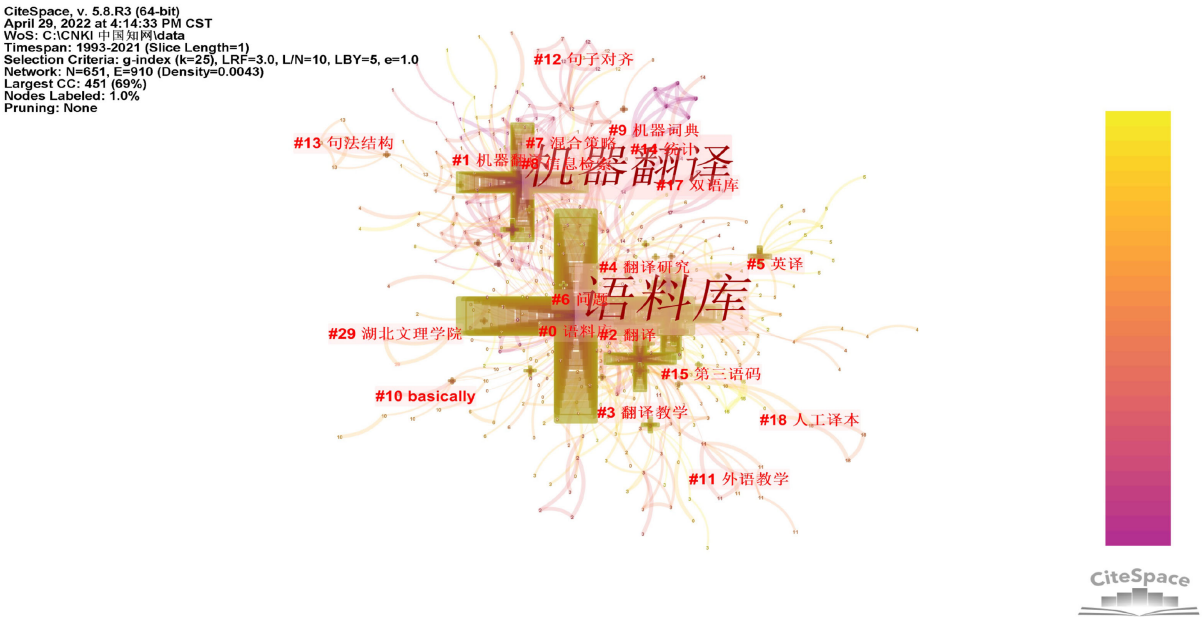

Figure 3. Keyword clustering map (LLR Algorithm)
图 3. 关键词聚类图谱(LLR 算法)

Table 2. Cluster Summary
表 2. 聚类汇总

聚类	文献量	文献起始年份	聚类内代表性关键词(LLR)
#0	153	2008	语料库(67.95, 1.0E−4)；机器翻译(25.59, 1.0E−4)；应用(15.62, 1.0E−4)；意义(8.9, 0.005)；翻译教学(7.99, 0.005)
#1	51	2009	机器翻译(64.56, 1.0E−4)；语块(16.22, 1.0E−4)；口语翻译(12.14, 0.001)；语料库(10.42, 0.005)；被动句(8.08, 0.005)
#2	45	2005	翻译(50.51, 1.0E−4)；显化(28.31, 1.0E−4)；人称代词(16.1, 1.0E−4)；语义韵(15.03, 0.001)；机器翻译(12.18, 0.001)
#3	42	2009	翻译教学(50.62, 1.0E−4)；翻译技术(14.94, 0.001)；术语翻译(14.94, 0.001)；实证研究(10.61, 0.005)；翻译过程(9.94, 0.005)
#4	42	2000	人工智能(41.65, 1.0E−4)；翻译词典(17.65, 1.0E−4)；语料库(12.99, 0.001)；命名实体(11.73, 0.001)；对齐模型(11.73, 0.001)
#5	41	2007	信息检索(20.72, 1.0E−4)；双语词典(20.72, 1.0E−4)；句子对齐(13.75, 0.001)；可比语料(6.84, 0.01)；平行语料(6.84, 0.01)
#6	39	2003	口译教学(35.18, 1.0E−4)；口译研究(29.23, 1.0E−4)；信息技术(23.31, 1.0E−4)；口译语料库(18.42, 1.0E−4)；多模态(11.58, 0.001)
#7	32	2008	混合策略(24.61, 1.0E−4)；基于统计(16.28, 1.0E−4)；基于实例(16.28, 1.0E−4)；基于规则(8.08, 0.005)；搭配抽取(8.08, 0.005)
#8	31	2012	翻译研究(23.6, 1.0E−4)；翻译批评(15.63, 1.0E−4)；文体(15.63, 1.0E−4)；wordsmith (7.77, 0.01)；译者(7.77, 0.01)

2.4. 研究热点

为了深入了解近 30 年国内语料库翻译学研究热点的变化情况，对 1993 年以来语料库翻译学研究热点的时区分布进行可视化分析，生成的知识图谱如图 4。

Figure 4. Key time zone map
图 4. 关键时区图

机器翻译自 1993 年至今一直是语料库翻译学应用领域的重要研究主题，随着平行语料库建设的发展，机器翻译过渡到基于语料库的现代机器翻译阶段。随着语料库基础工作的推进，2003 年平行语料库成为翻译研究中的热点，自此也成为该研究领域中最重要的一种语料库[11]。2010 年之后机器翻译研究融合了计算机科学、人工智能等其他学科知识。该研究热点在 2012 年出现了英译方向的迅猛增长，使中国文化走出去，利用平行语料库对比多版本的英译本，助推中国传统文化的对外传播。

3. 问题与展望

从上世纪九十年代开始，经过近三十年的发展，语料库翻译学进入快速发展阶段，从语料的构建，到研究方法、研究领域的不断拓展，再到现在发展成一种全新的研究范式，成为翻译学研究不可或缺的方向。不过，从以上分析可以看出，到目前为止，该领域的研究还存在一些问题：

第一：语料库设计和编撰首先要注意的是代表性、平衡性和规模大小。但对于这三个概念，则经常有人把三者混淆。盲目地把它们划等号[12]。语料库的代表性和平衡性是一个多维度筛选的过程，样本要注意随机抽样，语域分布要尽可能全面合理，要考虑各个语域的比例、词汇分布在语料库中是否合理，是否偏颇词汇或词语重复过多等诸多问题。

第二：目前建成的国内外语料库有笔语、口语占比不合理的问题。总体来说，书面语比重过大和口语部分转写耗时，一直是困扰语料库建设者的一个问题[13]。很多普通研究人员还是更倾向于建设书面语，因为这部分收集简单，但口语部分需要大量的人力进行语音听抄、转写，对转写人员的专业素质也要求较高，一般非一流的院校和普通学者很难承担这样的工作。

第三：语料库资源共享模式的探讨有待深入[14]。高昂的语料库版权费用限制了国内语料库研究的发展，为此国内外一些学者致力于开发一些免费的软件，如免费检索软件 Ant Conc，可以实现 Wordsmith 的主要功能。免费词性标注软件也有如 Brilltagger 和 Gotagger，但是正确率都不到百分之九十。既要考虑到语料库的版权问题，又兼顾降低门槛制作“在线语料库检索系统”是未来语料库发展的一个方向：

上海交通大学也把 Brown、LOB、CLEC、JDEST 四个语料库制作成了在线语料库检索平台。在线检索的优点在于不仅可以免费使用，而且保护了版权。但是问题在于如何优化检索速度、精度、稳定性和功能，使用户得到更好的体验[15]。

参考文献

[1] 胡开宝. 语料库翻译学概论[M]. 上海: 上海交通大学出版社, 2011: 3.

[2] 王克非. 语料库语言学探索[M]. 上海: 上海交通大学出版社, 2012.

[3] 王克非. 语料库翻译学——新研究范式[J]. 中国外语, 2006(3): 8-9.

[4] 陈悦, 陈超美, 刘则渊, 胡志刚, 王贤文. Cite Space 知识图谱的方法论功能[J]. 科学学研究, 2015, 33(2): 242-253.

[5] 廖七一. 语料库与翻译研究[J]. 外语教学与研究, 2000(5): 380-384.

[6] 刘康龙, 穆雷. 语料库语言学与翻译研究[J]. 中国翻译, 2006, 27(1): 59-64.

[7] 何春艳, 罗慧芳. 国内语料库翻译学研究动态的知识图谱分析(1993-2020) [J]. 中国科技翻译, 2020, 33(4): 17-20+42.

[8] 刘国兵, 常芳玲. 基于 CiteSpace 的国内语料库翻译学研究知识图谱分析[J]. 河南师范大学学报(自然科学版), 2018, 46(6): 111-120.

[9] 肖忠华, 戴光荣. 翻译教学与研究的新框架: 语料库翻译学综述[J]. 外语教学理论与实践, 2011(1): 8-15.

[10] 张继光. 国内语料库翻译学研究状况的科学知识图谱分析(1993-2014) [J]. 上海翻译, 2016(3): 34-40+61+93.

[11] 宋庆伟, 匡华, 吴建平. 国内语料库翻译学 20 年述评(1993-2012) [J]. 上海翻译, 2013(2): 25-29.

[12] 张新杰. 国内语料库语言学研究: 回顾与展望——基于核心期刊 24 年文献的统计分析[J]. 西安外国语大学学报, 2017, 25(2): 36-41.

[13] 王少爽, 高乾. 语料库翻译学的建构与拓展——王克非《语料库翻译学探索》评述[J]. 中国翻译, 2013, 34(2): 39-42.

[14] 于连江. 基于语料库的翻译教学研究[J]. 外语电化教学, 2004(2): 40-44.

[15] 王大鹏. 国内语料库发展现存问题与分析[J]. 渤海大学学报(哲学社会科学版), 2010, 32(3): 137-140.