

国内甲骨文研究热点、主题聚类及研究状况分析

——基于共词和社会网络分析视角

张瑞红

(安阳师范学院 计算机与信息工程学院;教育部甲骨文信息处理重点实验室;

河南省甲骨文信息处理重点实验室,河南 安阳 455000)

摘要:为客观反映我国近 20 年甲骨文领域的研究状况,以 CNKI 收录的甲骨文研究的期刊论文为研究对象,采用共词分析方法,从整体网、群体和成员 3 个维度分析共词网络,识别甲骨文研究热点;采用社会网络分析方法划分出 5 个凝聚子群,通过多维尺度分析验证子群的合理性,在该领域研究人员帮助下归纳研究主题,并运用战略坐标图分析研究主题的发展现状.研究发现,甲骨文研究热点词基本上与子群 2 重叠,子群 2 以古文字形体构造为研究主题,处于甲骨文研究领域的核心,研究趋于成熟,理论体系完善,为子群 1 的殷商甲骨文研究提供了方法和技术支持;甲骨文书法研究、考释和大辛庄甲骨文研究、古文字起源演变研究,这 3 个研究主题处于边缘且不太成熟.

关键词:甲骨文;主题聚类;研究状况;共词分析;社会网络分析

中图分类号:O29

文献标志码:A

从 1899 年王懿荣辨识出商代晚期龟甲兽骨上的文字开始,经过几代甲骨学人的不懈努力,甲骨文研究已经成为一门举世瞩目的国际性显学——甲骨学.漫漫两甲子,甲骨学者在甲骨文材料收集、整理、刊布方面付出巨大心血,为研究甲骨学奠定了坚实的基础.为了给甲骨文研究者提供参考,甲骨学者每隔一段时间就会对各个阶段的研究成果进行总结,1925 年王国维发表的《殷墟甲骨文字及书目》和 1926 年黄立猷发行的《甲骨类书目》是甲骨学论著目编著的标志^[1];1952 年胡厚宣编写的《五十年甲骨学论著目》出版,1999 年宋镇豪主编的《百年甲骨学论著目》出版,这是对甲骨学研究的专书、论文的汇编^[2].2019 年王子杨^[3]《近二十年来甲骨学研究述略》从甲骨资料著录、甲骨缀合、分类断代、甲骨字词考释、殷商史研究、甲骨工具书的编纂等方面,对 21 世纪近 20 年甲骨学研究成果进行了汇总.这些论著目提供了专书的书目信息,论文的题录信息,如果能对书目信息和论文题录信息进行文本挖掘,其意义不言而喻.本文以 2000—2019 年中国知网(CNKI)收录的甲骨文研究的期刊论文为研究对象,运用社会网络分析软件 Ucinet6.2,采用共词分析和社会网络凝聚子群分析的方法,从定量角度分析甲骨文研究的热点与主题,直观呈现该领域的研究状况,以期为未来研究甲骨文的人员提供一定的参考.

1 研究设计

1.1 研究方法和工具

关键词是作者对论文内容的提炼,高频关键词是某研究领域成员大量使用的词语,能在一定程度上表明该领域的研究热点.共词分析以文献集中共现的词对为统计对象,统计文献集的关键词之间在同篇文献中两两共现频次,构建由词对关联所形成的共词网络,关键词就是网络中的结点,网络中关键词间的远近可以反映主题内容的亲疏^[4].社会网络分析方法将网络中高频关键词间关联密切而聚合成的次级团体结构称之为凝聚子群,可以呈现网络中高频关键词的聚类,展现某领域的研究热点和主题结构^[5].

收稿日期:2020-07-13;**修回日期:**2020-10-15.

基金项目:国家社科基金重大委托项目(16@ZH017A3);教育部国家语委甲骨文等古文字研究与应用专项重点项目(YWZ-J023);教育部甲骨文信息处理重点实验室主任基金项目(OIP2019H005).

作者简介(通信作者):张瑞红(1979—),女,河南安阳人,安阳师范学院副教授,研究方向为甲骨文信息处理,E-mail: lxt262@163.com.

本文使用的高频关键词共现矩阵由文献题录信息统计分析工具 SATI3.2 构建,导入 Ucinet 将共现矩阵保存为可处理的格式,借助 Netdraw 可视化工具可生成高频关键词共现图谱,从整体网、群体和成员 3 个维度分析共词网结构特征,识别研究热点;采用社会网络分析的凝聚子群 Factions 方法,与该领域研究人员反复实验和讨论,划分出甲骨文研究的高频关键词凝聚子群,利用多维尺度分析验证子群划分的合理性;将 Netdraw 生成的凝聚子群图保存为 Pajek 格式,导入 VOSviewer 获得构建战略坐标图数据,分析各子群内部的强度和子群间的关联强度,解析每个研究主题的现状;发挥每个工具的优势,多层面、多维度展开分析。

1.2 数据来源

本文数据来源为 2000—2019 年间 CNKI 收录的期刊论文,为了确保获取数据的完整性,结合该领域研究人员建议和甲骨学相关著作选取了主题词,检索主题为“殷墟甲骨文”OR“甲骨文”OR“甲骨文字”OR“殷墟文字”OR“甲骨卜辞”OR“卜辞”OR“殷契”OR“刻辞”OR“契文”OR“龟甲文”OR“龟甲兽骨文”,检索时间为 2019 年 12 月 31 日,共获得 5 550 条记录,分年度进行严格筛选,筛除会议通知及纪要、展览信息、考古简报、书评、书序、书讯、人物传记、人物访谈、甲骨文书法作品和甲骨文公司相关的噪音文献,共获得检索记录 3 397 条。

1.3 研究数据整体分析

一个研究领域的整体发展状况、活跃程度和学术研究程度可从时间分布上得到反映,从 CNKI 收录的发文量统计图(图 1)可见,虽然随着时间的演进数据存在着波动与起伏,但是整体趋势是上升的。2014 年 5 月和 2016 年 5 月,习总书记在北京视察工作以及哲学社会科学工作座谈会上两次提到甲骨文,2019 年 11 月为“纪念甲骨文发现 120 周年座谈会”发来贺信,这些为以古文字为代表的‘绝学’、冷门学科注入新的研究动力,可以预期,未来甲骨文研究仍会呈现上升趋势。



图1 2000—2019年甲骨文研究发文量统计

Fig.1 Statistics of publication volume of Oracle Bone Inscriptions from 2000 to 2019

2 甲骨文研究热点分析

2.1 基于高频关键词的共现网络构建

2.1.1 高频关键词的提取

为了确保关键词的准确性,反映研究人员对特定关键词的共识度,消除人为因素的影响不对检索的主题词进行合并;只将简称、表达形式不一、写法不同的词语进行严格、审慎地预处理,如将《说文》、说文、《说文》、说文解字、《说文解字》统一为《说文解字》,将合集、《合集》、甲骨文合集与《甲骨文合集》统一为《甲骨文合集》;“殷墟”和“殷虚”经常在一篇论文的关键词中共现不进行统一合并。删除“比较”、“发展”、“先生”、“中华人民共和国”等无实际意义的词。

应用 SATI3.2 软件提取关键词,3 397 篇期刊论文,共 7 334 个关键词,累计总频次 15 779 次。为了尽可能看到近 20 年来甲骨文研究的全貌,参考《词频分析法中高频词阈值界定方法适用性的实证分析》^[6],根据研究需要将频次 12 次及以上,累计频次 4 885 次,占比约为 31% 的 96 个关键词确定为高频词,表 1 中是排在前 40 的高频关键词。

2.1.2 高频关键词共现矩阵的构造

应用软件 SATI3.2 构造一个 96×96 的共词矩阵,限于篇幅,摘取前 10 位高频词的共现矩阵,见表 2。表 2 中对角线是同行同列对应的关键词在文献集中出现的总次数,其余表格是行上对应的关键词与列上对应的关键词在文献集中共现的次数。

2.1.3 高频关键词共现图谱

高频关键词可以从一定程度上展示甲骨文研究的主要内容,却无法展示出关键词间的关联,共词分析通过关键词对共现的情况来展示关键词间的关联强度,从而展示出某学科或者某领域的研究热点。将 Ucinet 二值化处理过的高频关键词共现矩阵在 Netdraw 中呈现,根据度中心性控制关键词结点大小,结点越大表示关键词的度中心性越高,该关键词在网络结构中就越重要;关键词结点间的连线表示共现情况,连线粗细表示共现强度;从一个关键词结点发出的连线越多,表示该关键词是该领域研究的重点内容^[7]。高频关键词共现图谱,可见文中图 2。由图 2 可见,“甲骨文”、“殷墟文字”、“金文”、“文字”、“字形”、“甲骨文字”等结点大且周围连线较为密集,处于网络核心位置。

2.2 分析共词网络结构特征,识别甲骨文研究热点

从整体网、群体和成员3个维度分析共词网络特征,识别甲骨文研究热点,以期形成对甲骨文研究热点的可相互补充的完整论证。

表1 高频关键词表(前40个)

Tab.1 High-frequency keyword table(top 40)

高频词	词频	高频词	词频	高频词	词频	高频词	词频	高频词	词频
甲骨文	1 252	古文字	88	殷墟	71	甲骨学	46	会意字	35
殷墟文字	286	考释	86	甲骨文书法	67	殷墟卜辞	45	象形字	34
金文	187	本义	80	■	65	小篆	42	书法艺术	29
卜辞	128	《甲骨文合集》	78	文字	63	形声字	40	秦篆	29
商代	127	殷商	73	字形	59	祭祀	39	花东卜辞	28
《说文解字》	124	甲骨文字	73	缀合	58	篆书	38	王懿荣	27
甲骨卜辞	119	汉字	72	大篆	51	甲骨刻辞	37	许慎	27
甲骨	92	书体	71	殷墟甲骨文	47	书法	37	武丁	24

注:“■”是考释的甲骨字,无法正常显示的关键词。

表2 高频关键词共现矩阵(前10个)

Tab.2 High-frequency keywords Co-occurrence matrix(top 10)

	甲骨文	殷墟文字	金文	卜辞	商代	《说文解字》	甲骨卜辞	甲骨	古文字	考释
甲骨文	1 252	284	146	7	35	76	8	6	25	27
殷墟文字	284	286	47	0	0	39	6	5	16	0
金文	146	47	187	1	3	17	1	4	6	4
卜辞	7	0	1	128	18	0	0	12	1	3
商代	35	0	3	18	127	0	1	4	0	0
《说文解字》	76	39	17	0	0	124	2	0	5	2
甲骨卜辞	8	6	1	0	1	2	119	1	1	0
甲骨	6	5	4	12	4	0	1	92	1	3
古文字	25	16	6	1	0	5	1	1	88	33
考释	27	0	4	3	0	2	0	3	33	86

2.2.1 共词网络的整体网分析

使用网络密度、平均距离和聚类系数3个指标分析共词网络整体网状况。整体网密度取值范围 $[0,1]$,越接近1,说明关键词之间的联系越紧密,该网络对其中关键词的影响越大,关键词之间联系的越紧密。平均距离是连接任何两个关键词之间的最短途径的平均长度,平均距离短表示关键词间信息传播快,影响大,建立在“距离”基础上的凝聚力指数用来测量网络的整体凝聚力,取值范围 $[0,1]$,越接近1,网络越具凝聚力。聚类系数用来测度网络集团化程度,取值范围 $[0,1]$,越靠近1,网络越具凝聚力。关键词整体网络的特征值,见表3。

由表3可看到,网络密度为0.258,说明整体网络结构松散,关键词间共现的频次偏低,说明我国甲骨文研究的面较广,但是深度不够。平均路径1.745,距离基础上的凝聚力为0.629,距离是“1”和“2”的分别占比25.8%和73.9%,最大距离值是“3”,占比0.3%,说明绝大多数节点间的距离是2,两个节点间仅有一个中间节点,虽然网络密度不是太乐观,但是图2中孤立点不存在,且分散的点较少,所以0.598的整体网络聚类系数尚可,1.745的平均距离和0.598的聚类系数,说明关键词间具有一定的关系,存在高频关键词的网络小世界效应;从一定程度上也反映出甲骨文研究领域的关键词间具有关联性,网络连通性也较好。

2.2.2 共词网络中的群体——核心-半边缘-边缘分析

核心-半边缘-边缘结构指若干相互联系的节点形成一种中心紧密相连,外围分散稀疏的结构,可以用来分析哪些节点处

表3 关键词整体网络特征值

Tab.3 Overall network eigenvalues of keywords

项目	网络密度	平均距离	距离凝聚力	聚类系数
关键词	0.258	1.745	0.629	0.598

在社会网络的核心,哪些节点处在边缘的位置^[8].应用二值矩阵计算实际数据和理想模型间的相关系数为 0.508,取值范围为 $[-1,1]$,越接近 1,说明正相关的程度越大,0.508 的数值表明相关强度较大,说明共词网络是存在核心-半边缘-边缘结构的.核心-半边缘-边缘模型没有确定的划分方法,本文考虑关键词数量和核心度数值,确定一定取值区间得到,选择核心度数值大于等于 0.11 的为核心,大于等于 0.055 小于 0.11 的为半边缘,小于 0.055 的为边缘,共有 33 个词处于核心,见表 4.

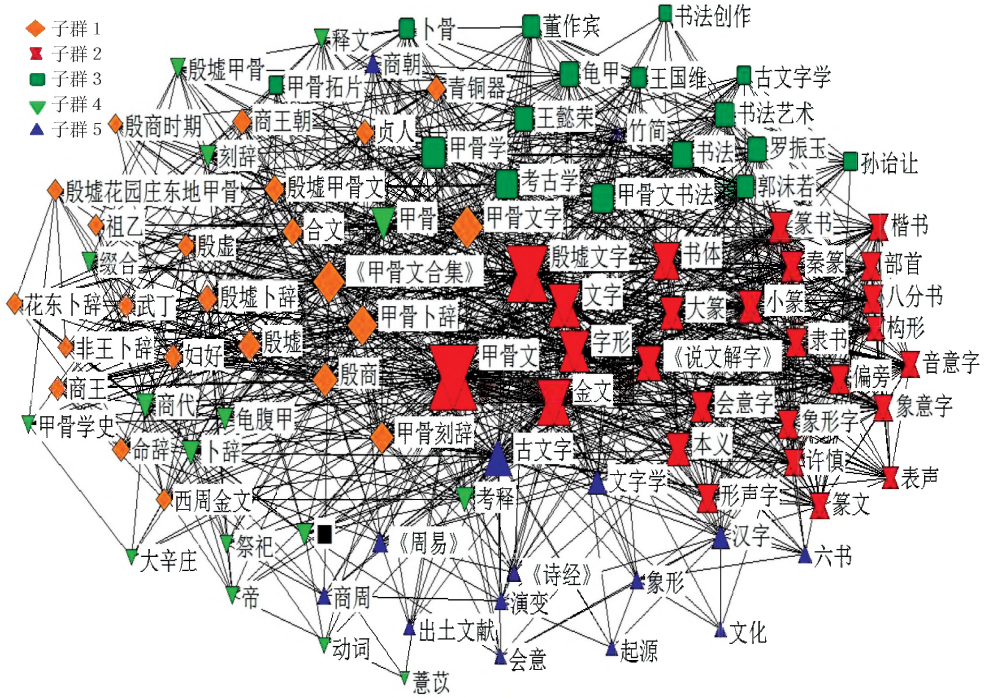


图2 高频关键词凝聚子群图谱

Fig.2 Map of condensed subgroup of high-frequency keywords

表 4 33 个核心关键词的核心度
 Tab.4 Core degree of 33 core keywords

关键词	核心度	关键词	核心度	关键词	核心度
甲骨文	0.269	甲骨卜辞	0.141	部首	0.121
殷墟文字	0.246	本义	0.141	象意字	0.120
金文	0.205	篆书	0.141	甲骨	0.119
文字	0.184	古文字	0.140	殷商	0.117
字形	0.184	秦篆	0.137	甲骨文书法	0.117
书体	0.168	会意字	0.133	甲骨学	0.117
甲骨文字	0.163	象形字	0.126	篆文	0.117
《甲骨文合集》	0.152	隶书	0.126	许慎	0.116
小篆	0.151	八分书	0.124	考古学	0.116
《说文解字》	0.147	偏旁	0.123	甲骨刻辞	0.115
大篆	0.147	形声字	0.122	音意字	0.115

2.2.3 共词网络中的成员——点的中心性分析

根据 LEE^[9]的研究,共词网络中结点的中心性可以用来衡量一个关键词在网络中的重要性,可以有效识别某研究领域的热点,主要有 3 个关键性指标:度中心性、中介中心性和接近中心性,应用二值矩阵计算 3 个中心性的值,本文仅列出 3 个中心性排在前 15 的关键词,如表 5 所示.

度中心性用来衡量在网络中哪些关键词最重要,“甲骨文”处在网络的核心位置,度中心性为 93,度中心性的平均值为 24.5,共有 41 个关键词大于平均值,这些关键词处在网络中较重要位置.中介中心性用来衡量哪些关键词具有桥梁作用,具有重要的信息传递和沟通作用;中介中心性的平均值为 35.375,共 19 个关键词大于平均值,这些关键词处在网络的核心,对其

他关键词有着较强的控制能力,其中“甲骨文”中介中心性为 897.068,具有最强影响力.接近中心性用距离来衡量关键词的中心程度,与其他关键词接近中心性值越大越不是网络的核心点,“甲骨文”接近中心性为 97 最小,与其他关键词之间的距离最小,在网络中处在核心位置,接近中心性的平均值为 165.75,共有 40 个关键词小于平均值,这些关键词与其他词之间联系较为紧密.由数据可以看到,度中心性和接近中心性的前 25 个关键词是相同的,除了表 5 中的还有“本义、殷商、篆书、考古学、殷墟、甲骨刻辞、甲骨文书法、秦篆、会意字、象形字”,中介中心性前 25 关键词除了“卜辞、殷墟卜辞、商代、合文”外都相同.因此,综合考量 3 个中心性,结合在一起的这 29 个关键词是研究热点词.对比表 4 和表 5 可见,核心度的排序与中心性的排序基本一致,处于核心位置的 33 个关键词,与共词网络成员分析得到的研究热点基本一致.

表 5 关键词的中心性数值(前 15)

Tab.5 The centrality value of keywords(top 15)

关键词	度中心性	关键词	中介中心性	关键词	接近中心性
甲骨文	93.000	甲骨文	897.068	甲骨文	97.000
殷墟文字	77.000	殷墟文字	370.049	殷墟文字	113.000
金文	60.000	金文	191.837	金文	130.000
文字	54.000	文字	151.911	文字	136.000
字形	52.000	字形	133.950	字形	138.000
甲骨文字	50.000	古文字	119.984	甲骨文字	140.000
《甲骨文合集》	50.000	甲骨卜辞	118.800	《甲骨文合集》	140.000
甲骨卜辞	47.000	甲骨文字	117.209	甲骨卜辞	143.000
古文字	45.000	《甲骨文合集》	117.190	古文字	145.000
书体	43.000	甲骨学	69.736	书体	147.000
《说文解字》	39.000	甲骨	58.629	《说文解字》	151.000
甲骨学	38.000	甲骨刻辞	47.723	甲骨学	152.000
大篆	38.000	《说文解字》	47.065	大篆	152.000
甲骨	37.000	殷商	46.806	甲骨	153.000
小篆	37.000	书体	46.316	小篆	153.000

综上所述,从整体网特征可见,近 20 年国内甲骨文研究存在热点;根据高频关键词词频、核心-半边缘-边缘分析和点的中心性分析,可以看到甲骨文研究的热点高频词基本上是一致的,以核心-半边缘-边缘分析的 33 关键词为参考,取每个分析角度的前 33 个,共有的关键词“甲骨学、殷商、甲骨、甲骨卜辞、甲骨刻辞、甲骨文、甲骨文字、殷墟文字、金文、文字、古文字、字形、《说文解字》、本义、会意字、象形字、书体、篆书、大篆、小篆、秦篆、甲骨文书法、《甲骨文合集》”是近 20 年该领域研究的热点词.这些热点词基本上与下文研究的子群 2 中的词一致.因此“子群 2 古文字形体构造”是甲骨文研究领域的热点和研究核心.

3 甲骨文研究主题的聚类分析

3.1 甲骨文研究的凝聚子群分析

共词聚类分析可反映关键词之间的亲疏关系,可将高频关键词共现网络关系转化为清晰的类群关系,通过聚类算法将关联密切的关键词聚合成类团,从而更好地展示学科领域研究的主题结构和主题内容^[16]. 社会网络分析中的凝聚子群分析,可以呈现网络中高频关键词的聚类.将高频关键词的二值化共现矩阵导入 Ucinet 中 Network 菜单下 Subgroups 的 Factions,依据不同子群划分出来的关键词,与该领域研究者反复论证,最终划分为 5 个子群,采用度中心性布局,得到国内甲骨文研究的高频关键词凝聚子群,如图 2 所示,图 2 中 5 个不同的图形代表不同的子群,不同的子群代表不同的聚类.

3.2 甲骨文研究的多维尺度分析

多维尺度分析将高维空间的数据变换成低维的数据,在低维空间中关键词用点表示,将高频关键词间的关联以平面距离的方式展示,平面中距离越接近说明关键词之间联系越紧,可以聚合成类团,可以更直观地展示关键词之间的关联^[17].将高频关键词二值矩阵导入 Ucinet 中 Tools 下 Scaling/Decompositong 的 Mctric MDS,根据相似性生成图 3 所示的高频关键词多维尺度图,关键词呈点状分布,每个关键词的位置可以呈现分析对象的相似性,距离越近的关键词之间关联就越大,高度相似的关键词聚在一起形成一个类团,位于中间的分析对象即为研究核心.从图 2 和图 3 可见,关键词凝聚子群图和多维尺度分析图的结果基本上一致,但是多维尺度分析图无法确定类团边界和数目,结合凝聚子群分析图可对多维尺度图中的研究类团进行划分.

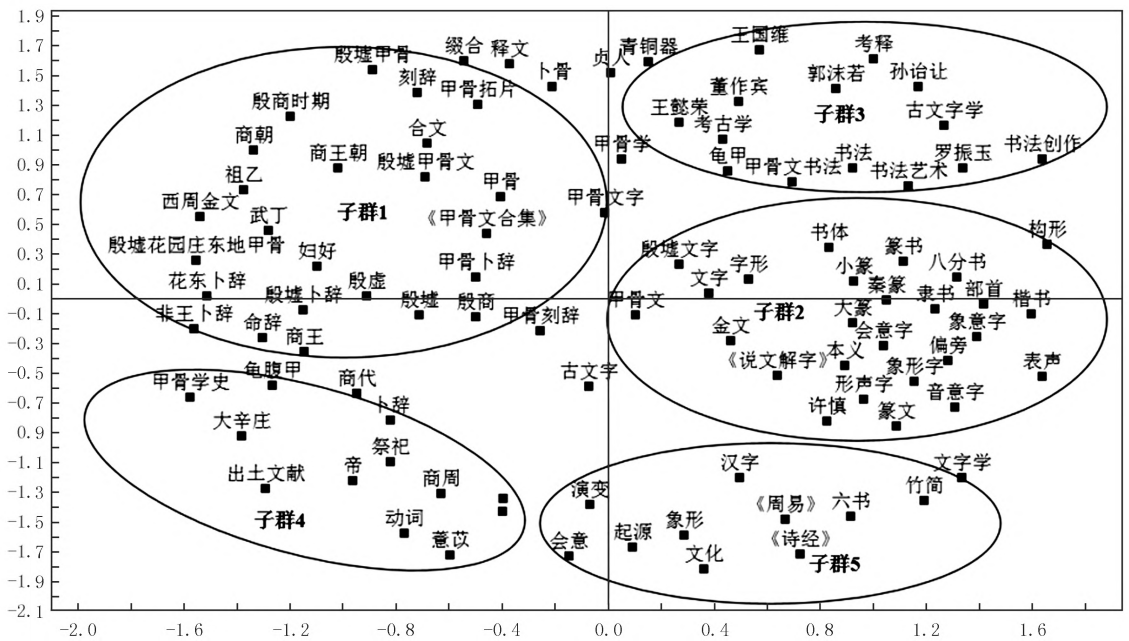


图3 甲骨文研究多维尺度分析图

Fig. 3 Multi-dimensional scale analysis diagram of Oracle Bone Inscriptions

3.3 甲骨文研究主题分析

结合高频关键词凝聚子群图和多维尺度图,对国内甲骨文研究的5个子群进行主题归纳和分析。

子群1:殷商甲骨文研究,包括殷商、商王、殷墟等23个关键词,研究主要集中在3个方面。一是殷商史研究,以甲骨卜辞材料为基础,综合铜器铭文等材料,从社会礼制、饮食文化、医疗生活等层面研究商王朝历史。二是“非王卜辞”研究,2003年《殷墟花园庄东地甲骨》出版后,掀起非王卜辞研究的热潮,在花东卜辞发现之前,非王卜辞比较罕见,花东甲骨出土极大丰富了非王卜辞的史料,对研究殷商时期家族祭祀等活动具有重要价值。三是字体特征为基础的甲骨文分组分类研究,如以“贞人”名字的字体研究分组等。

子群2:古文字形体构造研究,包括文字、字形、构形等26个关键词,古文字研究包括两个重要方面,一是古文字的形体结构,研究古文字要在综合考察形、音、义的前提下,以形体研究为中心,“寓义于形”是表意文字的特征;汉字结构有象形、指事、会意、形声、转注、假借的“六书”说法,并以六书分析汉字,汉字形体构造以象形字为基础,象形字又作为偏旁参与会意字、形声字的构造,古文字的形体结构是考释古文字的核心和关键。二是古文字的形体演变,古文字不断演化过程中,伴随着字体的变革,经历了甲骨文、金文、篆书(大篆、小篆)、隶书、楷书,其中小篆又名秦篆是过渡阶段,既是古文字的结束,又是今文字的开始。甲骨文发现之后,历史学家、古文字学家采用二重证据或者三重证据方法,将出土文献、古文字形体结构和传世文献结合,研究殷商史、考释古文字,借助古文字形体结构是考释古文字的有效方法,所以该主题一直是甲骨文研究的热点和重要内容。

子群3:甲骨文书法研究,包括甲骨文书法、罗振玉、董作宾等16个关键词,以“甲骨四堂”为代表的早期甲骨文研究者将甲骨文与书法艺术结合,董作宾更是将甲骨文引入书法篆刻。甲骨文书法研究体现在两方面^[12],一是“原生型”甲骨文书法研究,龟甲、卜骨是甲骨文的载体,对契刻在龟甲、卜骨上的甲骨文从刀法、线条条、形体、章法等方面研究其书法特征;二是“演进型”甲骨文书法研究,甲骨文发现后,以甲骨文为书体的书法创作,既有对早期甲骨学者甲骨文书法作品的述评与研究,也有对现当代甲骨文书法创作的研究。

子群4:考释和大辛庄甲骨文研究,包括“■”、大辛庄、龟腹甲等16个关键词。一是文字考释,“■”是考释的字,因无法输入用图片方式嵌入文中印刷,知网检索时关键词无法正常显示;近20年,考释非常典型的有“帝”字,以及从“词性”即甲骨文语法方面的研究。二是大辛庄甲骨文研究,2003年在殷墟之外首次发现和殷墟卜辞相似的商代卜辞,在甲骨学史上具有界标意义,发掘者缀合出大块龟腹甲,一些学者从不同的角度对大辛庄龟腹甲刻辞、卜辞进行了研究。

子群5:古文字起源演变研究,包括象形、会意、六书等15个关键词,甲骨文是迄今为止我国发现的年代最早的成熟文字系统,对于甲骨文之前的古文字起源,一些学者从不同的角度开展了一定的研究。另外,一些传世文献如《诗经》、《周易》等,记录了一些商周时期的传说,反映当时人类社会生产、生活实际等,为追溯古文字起源演变和考证商周历史提供佐证材料。

3.4 甲骨文研究主题的战略坐标图——研究状况分析

LAW等^[13]提出了战略坐标图,依此来刻画研究领域的主题聚类现状,反映主题聚类内部和聚类之间的关联状况。横轴代表向心性,反映主题聚类之间关联强度,值越大表明聚类之间关联越紧密,该聚类在其所属研究领域越为核心,越是该领域研

究者关注的焦点,在该领域发挥越重要的作用;纵轴代表密度,值反映主题聚类内部关联强度,值越大表明聚类内部关联越强,该聚类在其所属研究领域研究时间越长,研究者较多且已形成较完善的理论体系,越趋于成熟,其维持和发展的能力越强.

将 Netdraw 中生成的凝聚子群图保存为 Pajek.net, Pajek.clu 和 Pajek.vec 格式,在 VOSviewer 中以上述 Pajek 的 3 种方式打开,图谱呈现后,不做处理,直接保存后即可获得聚类、标签的信息和连线的数据,复制到 Excel 中采用总和平均值方法,计算每个凝聚子群的向心度值和密度值,向心度为子群内所有结点与其他子群内所有结点连线总和的平均值,密度值为子群内所有结点之间连线总和的平均值,向心度值和密度值见表 6,向心度和密度的平均值取整,即坐标原点(470,22),战略坐标图见图 4.

表 6 向心度和密度的数值

凝聚子群	向心度(X 值)	密度值(Y 值)
子群 1	563.500	24.391
子群 2	637.000	37.385
子群 3	392.000	22.625
子群 4	392.000	15.500
子群 5	367.500	13.933

第一象限包括两个子群,说明子群 2 和子群 1 是甲骨文研究领域的焦点和重点.子群 2 对应的古文字形体构造研究,向心度和密度最高,说明子群 2 与其他子群之间的关联紧密,在其所属研究领域

处于核心,是该领域研究者关注的焦点,在该领域具有重要作用;同时该子群内部关联强度大,在其所属的甲骨文研究领域研究时间较长,研究人数多并已形成较为完善的理论体系,研究趋于成熟,在该领域自身维持和发展的能力很强.子群 1 对应的殷商甲骨文研究,仅次于子群 2,同样是该领域中研究的重点,与其他子群之间联系较为广泛,但是子群内联系没有子群 2 密切,说明相较于子群 2 研究成熟度偏低些,研究人数是古文字研究的一个子群.

第二象限子群 3 对应的甲骨文书法研究,向心度低于平均值,密度接近平均值,在甲骨文研究领域里研究的人比较多,但是处于比较边缘的研究主题,因其处在第二象限和第三象限的交界,所以研究还不是太成熟.

第三象限包括子群 4 和子群 5,这两个子群的向心度和密度均低于平均值,说明与其他子群联系不太紧密,学界没有给予广泛关注,两个子群内部的研究松散,在甲骨文研究领域不是重点研究内容,是不成熟的边缘研究主题.

4 结 论

文章采用共词和社会网络分析方法,利用共词网结构特征,识别研究热点,采用社会网络分析方法划分出 5 个凝聚子群,通过多维尺度分析验证子群合理性,归纳研究主题,并运用战略坐标图分析研究主题的发展现状.主要研究结论如下.

(1)利用共词网络结构特征,从整体网、群体和成员 3 个维度识别的研究热点与高频关键词基本上一致,而且这些热点词与子群 2 中的词基本上一致,根据战略坐标图分析可知,子群 2 的古文字形体构造研究处于甲骨文研究的核心,研究趋于成熟,理论体系完善.

(2)甲骨文领域的 96 个高频关键词,被划分为 5 个凝聚子群,运用多维尺度方法对凝聚子群划分的合理性进行了验证,结合两者归纳出 5 个研究主题.该领域研究人员认为子群 4 也可合并到子群 1,子群 5 也可合并到子群 2,但是社会网络分析软件划分群时有一定的错误率,如果划分为 3 类,错误率太高,反而不利于归纳分析,故与该领域研究人员反复讨论后,将凝聚子群划分为 5 类,在多维尺度图上关键词呈现的状况与凝聚子群基本上一致,说明本文 5 个研究主题的划分是合理的,研读该领域学者的典型文献,对每个研究主题进行归纳.

(3)甲骨文研究领域的 5 大主题,从战略坐标图看,古文字形体构造研究处于核心,实际上为殷商甲骨文研究提供了方法和技术,就甲骨文而言子群 1 的殷商甲骨文研究是核心,是古文字研究中甲骨文学者关注的重点.子群 3 甲骨文书法研究人数虽多,但对于甲骨文研究来说是比较边缘的主题,成熟度偏低.子群 4 和子群 5 的研究比较零散,不是该领域关注的主要内容.

近 20 年甲骨学人利用新出土的材料和缀合恢复的重要甲骨材料,不断丰富和深化着甲骨文的研究,让甲骨学的国际显学地位不断得到彰显.本文在该领域研究人员指导下,从定量角度对甲骨文研究热点和主题的归纳具有一定的代表性,但是在数据处理和主题聚类分析中存在着一定的主观性,希望可为甲骨文研究提供一定参考.

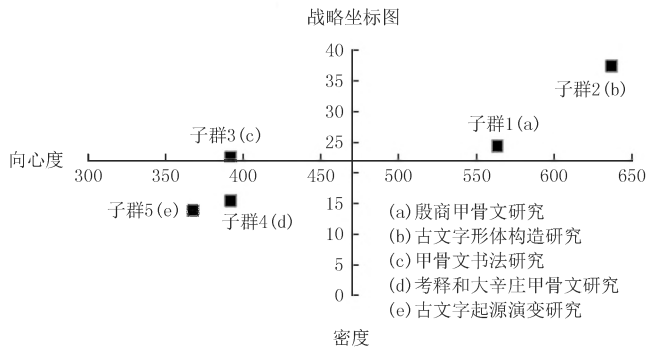


图4 甲骨文研究主题战略坐标图

Fig. 4 Strategic coordinate chart of Oracle Bone Inscriptions research themes

[1] 郑飞洲.甲骨学研究的基石和起点:评胡厚宣的《五十年甲骨学论著目》[J].书品,2002(2):19-25.

- ZHENG F Z.The Base and Starting Point of the Study of the Oracle Bone Inscriptions;A Review of Hu Houxuan's Book Fifty Years of Works on Oracle Bone Inscriptions[J].Shupin,2002(2):19-25.
- [2] 王宇信.甲骨学研究一百年[J].殷都学刊,1999(2):1-14.
- WANG Y X.One Hundred Years of Research on Oracle Bone Inscriptions[J].Yindu Academic Journal,1999(2):1-14.
- [3] 王子杨.近二十年来甲骨学研究述略[J].中国书法,2019(12):69-178.
- WANG Z Y.A Brief Account of the Research on Oracle Bone Inscriptions in Recent 20 Years[J].Chinese Calligraphy,2019(12):69-178.
- [4] LE M M,COURTIALI J P,SENKOVSKA E D,et al.The Dynamics of Research in the Psychology of Work from 1973 to 1987;from the Study of Companies to the Study of Professions[J].Scientometrics,1991,21(1):69-86.
- [5] MA H W,ZENG A P.The Connectivity Structure,Giant Strong Component and Centrality of Metallic Networks[J].Bioinformatics,2003,19(11):1423-1430.
- [6] 刘奕杉,王玉琳,李明鑫.词频分析法中高频词阈值界定方法适用性的实证分析[J].数字图书馆论坛,2017(9):42-49.
- LIU Y S,WANG Y L,LI M X.An Empirical Analysis of the Applicability of the Threshold of High Frequency Words in Word Frequency Analysis[J].Digital Library Forum,2017(9):42-49.
- [7] 刘启元,叶鹰.文献题录信息挖掘技术方法及其软件 SATI 的实现;以中外图书情报学为例[J].信息资源管理学报,2012(1):50-58.
- LIU Q Y, YE Y.The Technology and Method of Bibliographic Information Mining and the Realization of Its software SATI;A Case Study of Chinese and Foreign Library and Information Science[J].Journal of Information Resource Management,2012(1):50-58.
- [8] BORGATTI S P,EVERETT M G.Models of Core Periphery Structures[J].Social Networks,1999,21(4):375-395.
- [9] LEE W H.How to Identify Emerging Research Fields Using Scientometrics:An Example in the Field of Information Security[J].Scientometrics,2008,57(3):357-377.
- [10] 赵宾,董颖,杨晓杰.国内信息生态研究的图谱与热点主题;基于文献计量学共词分析的视角[J].情报科学,2017,35(9):62-66.
- ZHAO B,DONG Y,YANG X J.Atlas and Hot Topics of Information Ecology Research in China;Based on the Perspective of Bibliometrics Co-word Analysis[J].Information Science,2017,35(9):62-66.
- [11] 张文彤.SPSS II 统计分析教程[M].北京:北京希望电子出版社,2002:123.
- [12] 陈爱民.从宗教到艺术:甲骨文书法的观念转换[J].艺术家,2019(3):145-147.
- CHEN A M.From Religion to Art;Concept Transformation of Oracle Bone Inscription Calligraphy[J].Artist,2019(3):145-147.
- [13] LAW J,BAUIN S,COURTIALI J,et al.Policy and the Mapping of Scientific Change:A Co-word Analysis of Research into Environmental Acidification[J].Scientometrics,1988,13(3/4):251-264.

Hotspots, theme clustering and research status of Oracle Bone Inscriptions in China

—From the perspective of Co-word and social network analysis

Zhang Ruihong

(School of Computer and Information Engineering; Ministry of Education Key Laboratory of Oracle Bone Inscriptions Information Processing; Henan Key Laboratory of Oracle Bone Inscriptions Information Processing, Anyang Normal University, Anyang 455000, China)

Abstract: In order to objectively reflect the research status of Oracle Bone Inscriptions in the 20th century in China, this paper takes the journal papers of Oracle Bone Inscriptions in CNKI as the research object. The method of Co-word analysis is adopted to analyze the Co-word network from the three dimensions of the whole network, the group and the members. It identifies the research hotspots of the oracle bone inscriptions. Five cohesive subgroups are divided by social network analysis method, and the rationality of the subgroups is verified by multidimensional scale map analysis. The research theme is summarized with the help of researchers in the field, and the development status of the research theme was analyzed by using strategic coordinate map. The hotspot words in the study of oracle bone inscriptions are basically overlapped with subgroup 2, which takes the ancient character form structure as the research theme and is at the core of the research field of oracle bone inscriptions. The research tends to be mature and the theoretical system is perfect, which provides the method and technical support for the study of oracle bone inscriptions during the period of Yin and Shang in subgroup 1. The three themes of the study of oraclebone inscriptions calligraphy, the study of interpretation and the study of Daxinzhuang oracle bone inscriptions, and the study of the origin and evolution of ancient characters are marginal and immature.

Keywords: Oracle Bone Inscriptions; theme clustering; research status; co-word analysis; social network analysis