

新文科背景下的日语语言文学硕士论文 计量文献学考察 (2012—2022)

刘善钰, 王子睿

[摘要] 使用文献计量学方法和编程软件 R 语言, 针对中国知网 2012—2022 年登载的日语语言文学专业硕士毕业论文进行文献量化分析发现, 近 11 年来, 日语语言文学专业的硕士毕业论文总体数量呈现减少趋势, 其中外国语高校贡献数量逐年增多, 占比逐渐增大; 对应分析显示, 各大高校结合本校日语学科的优势呈现出不同的研究特点, 湖南大学、上海外国语大学等高校在语料库语言学研究方面表现突出, 而部分高校仍偏重传统文学研究。主题分类模型的结果显示日语语言文学专业的毕业论文仍然以传统人文研究为主, 缺乏对日本社会、国别研究等领域的关注。新文科背景下的日语语言文学专业硕士论文研究应当积极尝试与自然科学学科之间的交叉与融合, 使用跨学科方法, 结合跨学科知识进行研究, 以推动日语学科的长足发展。

[关键词] 量化分析; 日语语言文学; 计量文献学; R 语言

[中图分类号] H36

[文献标识码] A

引言

2019 年教育部召开“六卓越一拔尖”计划 2.0 启动大会, 正式拉开了新文科建设的序幕。新文科是指对传统文科进行学科重组, 实现文科内部以及文科与自然科学学科之间交叉与融合之后形成的文科。^[1] 新时代的文科发展要求外语研究人员摒弃学科本位主义, 使用跨学科思维与方法解决实际问题。硕士研究方向一定程度上体现了某一院校该学科的研究方向。通过对日语语言文学专业 11 年来的硕士毕业论文进行量化研究, 可以对该学科各院校的研究特点及强势研究领域进行描述和分析。在新文科建设的大背景下, 研究日语语言文学专业学

科的量化规律, 有助于掌握当前研究生的研究方向, 以利于观察硕士论文研究方向与院校之间的联系, 为新文科建设提供可观察的真实数据支撑。

日语语言文学专业 (学科代码: 050205), 一般被称为日语专业学术型硕士 (以下简称日语学硕), 是外国语言文学的二级学科。日语笔译和口译专业 (学科代码: 055105), 一般被称为日语专业的专业型硕士 (以下简称日语专硕), 是翻译硕士专业的二级学科。日语学硕和专硕是日语专业本科生报考硕士专业的主要方向。前者主要从事日语语言学研究、日本文学研究、日本文化研究, 后者主要从事日语的口、笔译实践以及翻译学研究。有关日语专业硕士论文的前沿研究中, 闫鑫基于问卷调查法和

[收稿日期] 2023-05-23

[基金项目] 2021 年广西高等教育本科教学改革工程项目“基于成果导向的《高级日语》课程思政教学改革与实践”(2021JGB125)

[作者简介] 刘善钰, 广西大学外国语学院讲师, 博士, 研究方向: 日语教育、日语语言文学、语料库语言学; 王子睿, 北京语言大学博士研究生, 研究方向: 基于计量方法的文学、翻译研究。

因子分析等方法对 100 名日语专业硕士生的毕业论文选题方向进行了考察,指出当前研究生硕士论文课题的制定 36% 是由个人兴趣爱好来决定,日语语言文学专业的研究生在制定论文题目时,存在专业知识不足和学术能力不够的特点。^[2]鉴于当前日语语言文学专业亟须在新文科背景下进行适当的学科调整,以重新审视当前学科名与实的问题,有必要针对日语语言文学专业硕士毕业论文数据进行分析研究。本研究使用 R 语言工具对 2012 年到 2022 年 11 年间的日语学硕毕业论文进行文本挖掘与量化分析,目的主要有二:第一,了解学科发展动向、为日语语言文学专业的研究生提供选题思路;第二,了解不同高校的热点研究方向,为该学科研究生阶段的高校教育进行新文科改革提供实证数据。

一、文献计量研究与研究设计

(一) 文献计量研究

文献计量法(Bibliometrics)是一种用于描述和分析某一学科或研究领域的动态与进展的计量方法,在现代计算机技术的辅助下,可以用清晰明了的知识图谱来可视化文献分析的结果。文献计量法作为一种定量分析方法,以科技文献的各种外部特征作为研究对象,采用数学与统计学方法来描述、评价和预测科学技术现状与发展趋势。^[3]文献计量学的可视化图可以方便地进行数据的解释与说明,对研究对象进行详尽的统计与分析,同时有助于挖掘信息之间的内在联系。

(二) 研究设计

在文献计量分析中常见的文献计量分析工具有 citespace、vosviewer、cooc 等软件,本研究使用编程软件 R 语言进行数据处理以及可视化。R 语言是一个自由、免费、源代码开放的软件,其拥有便利、快捷的数据处理与较强的可视化能力。在 R 语言编程与可视化技术的辅助下,可以用清晰明了的知识图谱来可视化文献分析的结果。

在使用数据方面,本研究主要通过中国知网抓取了 2012—2022 年的日语语言文学专业硕士毕业

论文文献数据,去除了部分噪声数据,保留了包括毕业年份、毕业学校、论文题目、论文摘要、论文关键词等信息。运用文献计量学的方法对以上信息进行多维度考察,例如不同年份学硕的贡献数量、不同高校的贡献数量、不同高校的研究方向等等。但由于部分高校的硕士论文不收录在中国知网数据库中,且查找困难,本研究不将其作为数据,只将知网收录的 4 699 条数据集(检索时间节点为 2022 年 12 月 30 日)作为本研究的观测数据。

在工具方面,选取 R 语言的 tibble 包、ggplot 包、quanteda 包、stringr 包、jiebaR 包作为数据处理工具,对数据进行结构化处理,基于词袋模型(bag of words)对文本数据进行可视化分析,以高校为单位计算 10 年来所有毕业论文的关键词和摘要文本。

在统计方法上,本研究除了描述统计,还使用了无监督学习算法中的对应分析法(Correspondence Analysis,简称 CA 分析)与隐含狄利克雷分布主题模型(Latent Dirichlet Allocation,简称 LDA 模型)。无监督学习最大的特点即不需要手工标注的训练集,仅需要文本文档以及指定主题或者聚类的数量 k 就可以对大规模的数据进行分类、降维。对应分析法的主要思想在于揭示同一变量的各个类别之间的差异,以及不同变量各个类别之间的对应关系。本研究利用对应分析法考察不同高校和不同学科关键词之间的对应关系,LDA 模型可以分析所选取论文摘要中所隐含的主题,帮助理解学科研究的内在主题。

二、数据分析

(一) 文献数量年度分析

为了了解不同年份的论文产出情况,对每个年份的日语语言文学专业硕士论文数量进行了统计分析,用 ggplot2 包绘制点图与拟合线(见图 1)。观察图 1 可知在过去的 11 年里,日语语言文学专业的毕业论文数量在总体上呈现逐年减少的趋势,尤其在 2021 年数量降到最低,只有 255 篇,可以推测 2018 年的招生数量明显低于其他年度,或由于

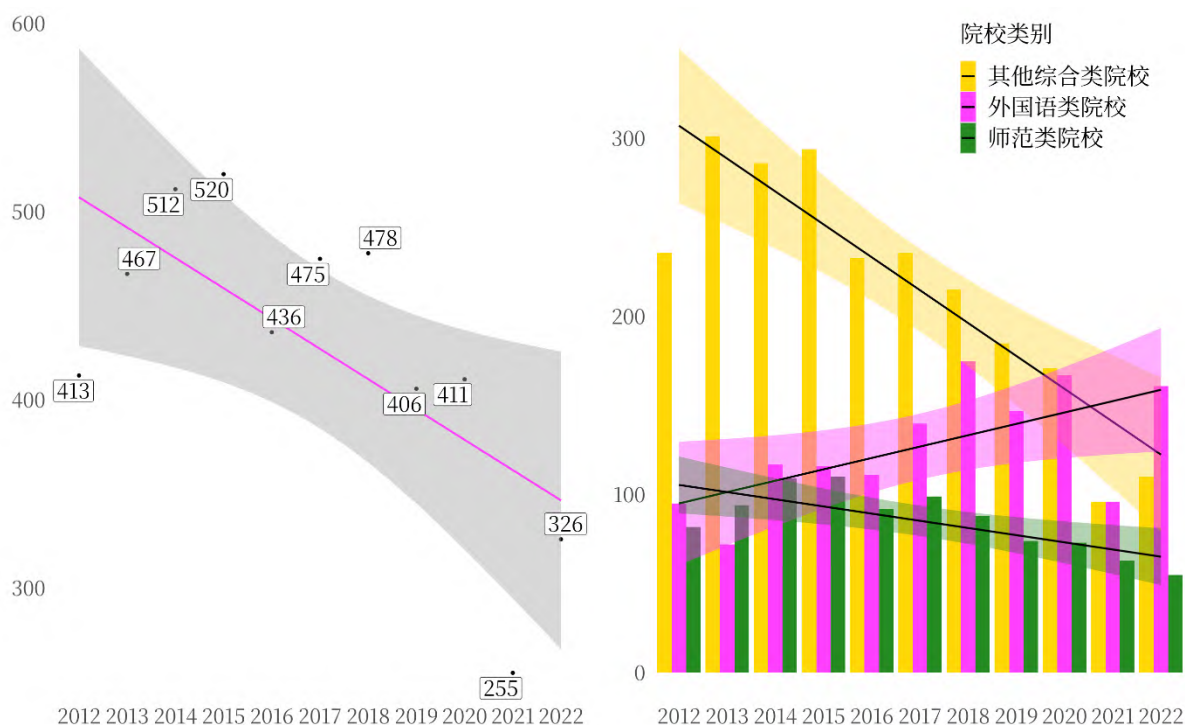


图 1 2012—2022 各年份论文贡献数量

新冠疫情等因素影响导致无法按时完成毕业论文的人数增加。将数据按照院校类别变量进行分类,即分为综合类院校、师范类院校、外国语类院校来观察,可以看出随着年度的增加,综合类院校及师范类院校的贡献数量减少,而外国语类院校贡献数量呈现增加的趋势。

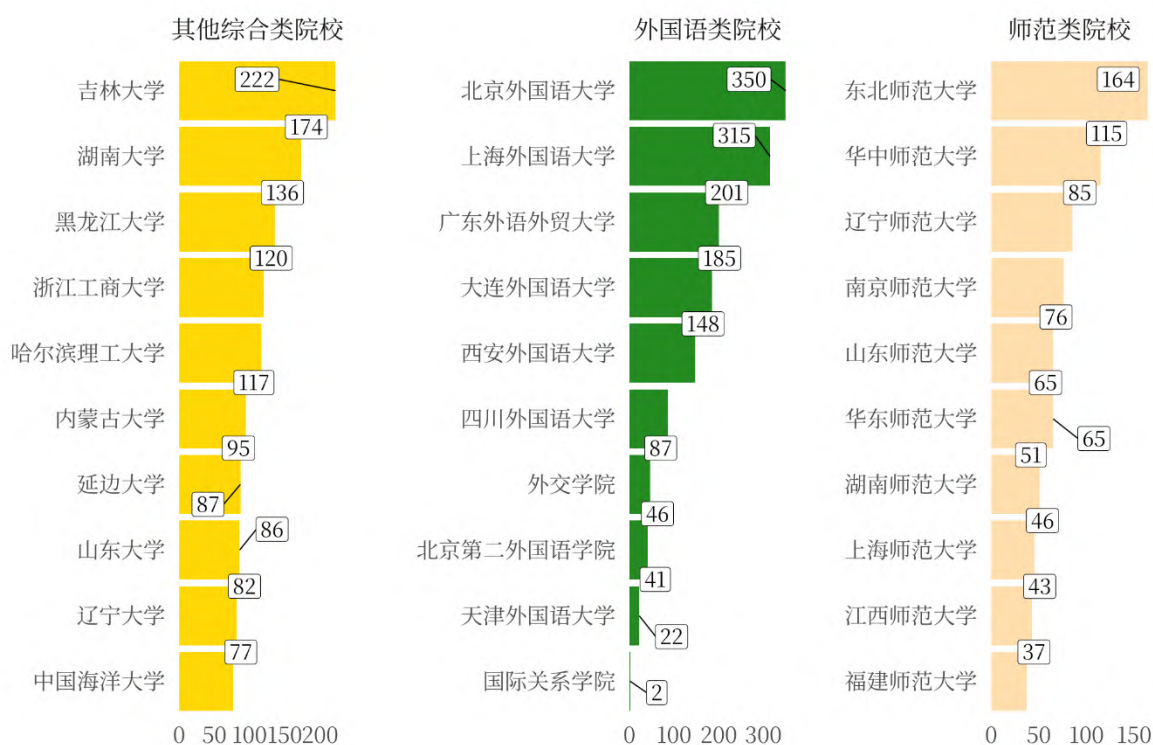
(二) 论文贡献单位分析

通过统计综合类院校、师范类院校、外国语类院校三类院校的论文贡献数量得知,11年里日语语言文学专业硕士论文的贡献数量在不同类别高校占比不同,由低到高依次为师范类院校 19.98%、外国语类院校 29.73%、综合类院校 50.29%。进一步观察各院校的具体论文贡献数量(见图 2)发现,在综合类院校中吉林大学、湖南大学、黑龙江大学贡献数量排列前三,浙江工商大学和哈尔滨理工大学虽为非文科类院校但论文贡献数量也较多,位列第四、第五。在外国语类院校中,北京外国语大学、上海外国语大学、广东外语外贸大学的贡献数量排列前三。在师范类院校中,东北师范大学、华中师范大学、辽宁师范大学的贡献数量排列前三,但总

体数量低于综合类院校和外国语类院校。

(三) 学科研究领域分析

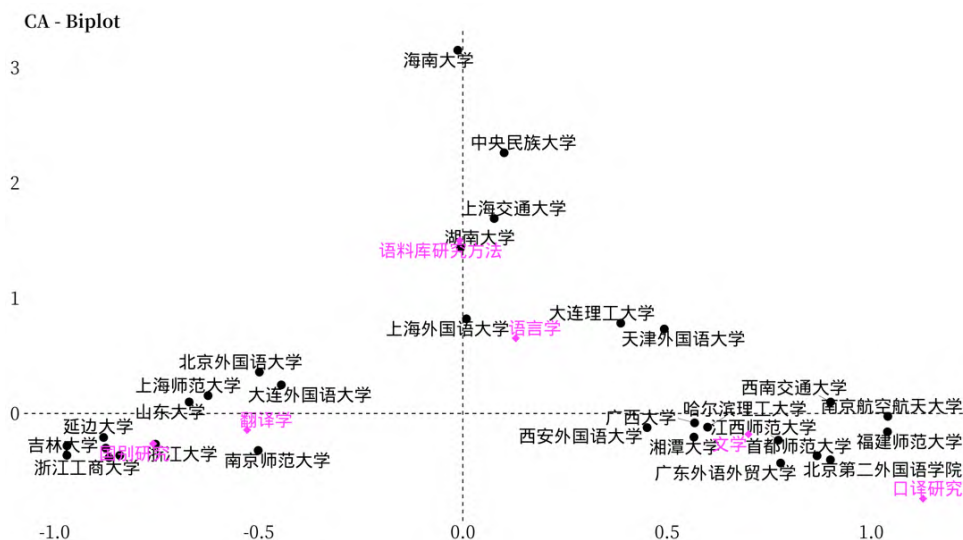
语言类研究生的研究方向大多集中在语言学、文学、翻译学等人文社科领域,又因研究方法、研究对象、研究工具等的不同,上述的学科领域下又可细分为不同的研究方向。例如语言学中的认知语言学与语料库语言学,文学中的伦理学、比较文学等,翻译学下又可以细分为口译、笔译研究,此外还涉及文化、国别研究、国际关系研究等等。本研究通过正则表达式抽取了该学科硕士论文的关键词并按研究领域对论文进行分类,例如包含“二语习得”“偏误”“助词”等语言学常见词语的列为语言学研究领域内的论文,包含“译本”“译者”的研究归为翻译类研究,包含“口译”“同声传译”等的将其视为口译研究,包含“小说”“文学作品”“变异学”“比较文学”等字符的研究划为文学类研究。此外,本文研究还抽取了语料库工具类的关键词,观察不同学科研究之间以及不同高校的日语语言文学学硕论文中语料库研究方法以及工具的使用情况。



在数据抽取方面,本研究将每一篇论文的摘要作为一条观测数据,院校、论文年度等其他因素作为变量。如某论文摘要中有“译本”二字,则此篇毕业论文则被认定为翻译研究领域的论文,以此类推计算所有研究领域内的论文数量。本研究选取了对应分析(CA分析)作为多元统计手法。作为一种视觉化的数据分析方法,对应分析能够将不同学

科关键词作为变量，将学校单位作为观测值，观察关键词与学校单位之间乃至关键词之间、学校单位之间的相关关系。

如图 3 所示(因高校数量太多, 仅选取累计解释度最高的前 30 所高校展示), 语言学相关的词语与语料库研究方法相近, 说明语言学研究领域的论文与语料库研究工具相关性更高, 互动性更强, 但



也说明了语料库技术在其他研究领域的应用研究还不够深。例如文学、口译等研究领域, 这些研究往往以内省式的传统研究方法为主, 缺乏基于定量数据的分析。通过图 3 可以看到湖南大学日语学硕的毕业论文与语料库研究方法分布在非常近的地方, 其周围上海外国语大学、上海交通大学在此类研究中也比较突出。另外, 文学研究领域方面, 周围聚集的高校比较多, 说明这些高校的日语语言文学专业硕士生的研究方向还是以文学研究为主。此外, 北京第二外国语学院在口译研究方面较为突出, 浙江大学与浙江工商大学等高校的研究方向多集中在国别研究方面。

图 4 为毕业论文信息中包含“语料库”一词的论文数量统计图, 从中可以发现上海外国语大学、大连外国语大学、北京外国语大学、湖南大学、吉林大学等高校使用语料库的论文数量较多, 其中上海外国语大学在语料库使用方面格外突出, 共计有 64 篇硕士论文使用了语料库技术。

(四) 文献关键词分析

毕业论文关键词是论文主体内容的高度凝练,

本研究选取了毕业论文关键词中出现频率最高的 50 词, 构建词袋模型与特征矩阵, 构建固定上下文窗口的共现矩阵 (Co-Occurrence Matrix with a fixed context window)。例如, “谷崎润一郎”“中国”作为一组关键词在一篇论文中出现, 共计词数为 1, 但是在别的论文中可能是“谷崎润一郎”“唯美主义”, 那么就可以将“谷崎润一郎”作为中心向其他两个点画线, 某词出现的频率越高, 这个词的中心性就越强。由图 5 可以清晰地看到不同关键词之间的联系与出现频率, 例如“正用”属于二语习得研究领域的词语, 该词与“中国人日语学习者”“问卷调查”共同出现的频率较高, 所以分布在相近的位置。高频出现的作家名字有“夏目漱石”“三岛由纪夫”“松本清张”“谷崎润一郎”“吉本芭娜娜”等, 说明在日语语言文学专业的论文中多以上述的文学作家为研究对象, 其中比较研究、女性观、女性意识、战争观是此类论文主要的研究视角 (见图 5)。

文学、翻译学研究的关键词往往会注明研究的作品, 即文学作品名。本研究收集的硕士论文关键

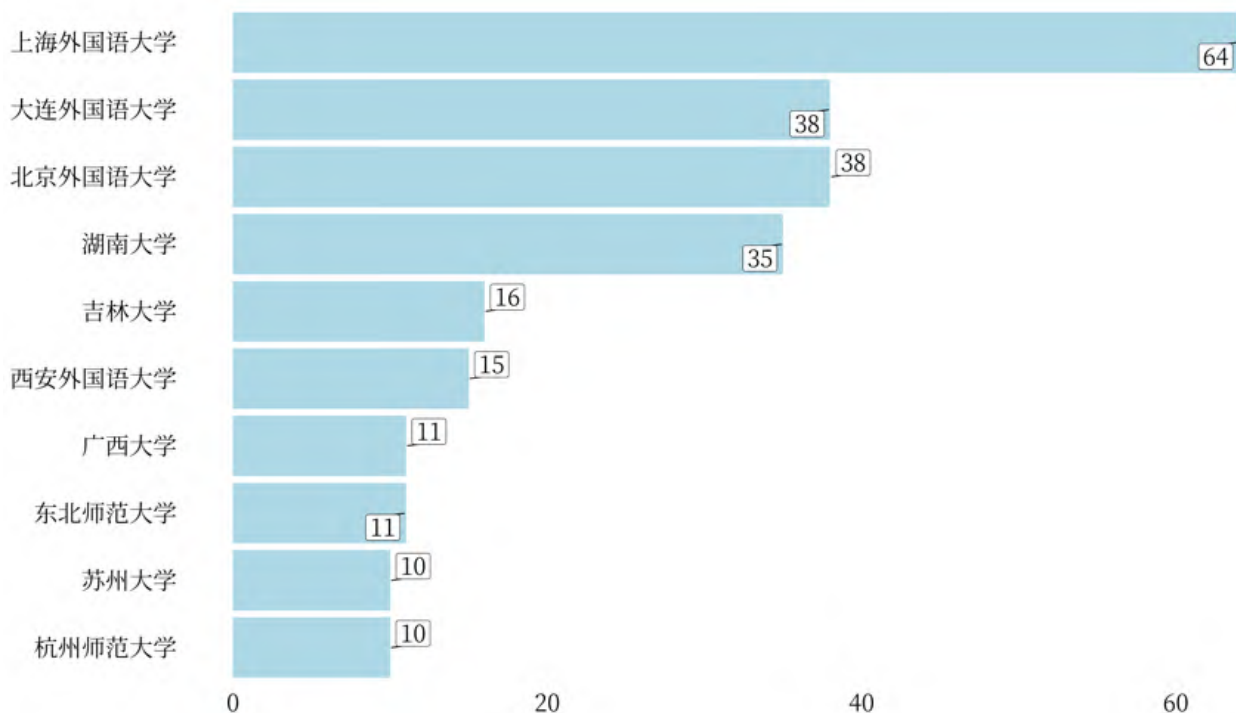


图 4 “语料库” 相关论文数 (以高校为单位)



图 5 关键词共现分析图

词中含有书名号的有 804 条, 约占比 17.1%, 说明以文学作品为研究对象的硕士论文占比不少。用正则表达式检索所有论文中带有书名号的关键词, 并对其进行频数统计, 导入 WordCloud2 包进行词云图绘制得到图 6。观察图 6 可以看到位于词云图中心的有日本文学经典古典名著《源氏物语》、日本作家村上春树的长篇小说《1q84》、川端康成的中篇小说

《雪国》，这些作品名出现次数均大于 10 次，日语语言文学专业研究生通过多个角度对这些文学作品进行了探讨。此外还有中国古典文学作品例如《红楼梦》《聊斋志异》等的日译也被反复分析研究。

此外,本研究还抽取了硕士论文关键词中出现的文学作家名,制作成词云图,如图7所示。其中关键词中太宰治的出现频率为50次,是目前日



图 6 研究对象文学作品词云图

语语言文学专业研究生文学研究中最常见的研究对象, 其次是夏目漱石 38 次、川端康成 30 次、井上靖 25 次, 安部公房 21 次, 说明这些作家是日语专业硕士论文中常见的研究对象。

(五) 文献摘要分析

使用隐含狄利克雷分布主题分类模型 (LDA) 来分析日语语言文学学硕论文摘要中所隐含的主题, 主要方法是: 在使用 jiebaR 分词时, 将论文中的关键词作为用户词典导入, 提高摘要部分的分词精度, 再对论文摘要进行分词后使用 LDA 模型将每篇文档的主题按照概率分布的形式算出, 结果如

图 8 所示。可将日语学硕的毕业论文主题归纳为以下五大类: 中日政治经济社会研究、中日思想发展史研究、对照语言学与翻译学研究、文学与文学意象研究、二语习得与应用语言学研究。这说明日语语言文学专业的毕业论文多以传统文科研究主题为主, 主要涉及社会文化、文学与翻译、语言学与应用语言学领域的内容。

三、结果与讨论

(一) 结果分析



图 7 以作家为研究对象的词云图

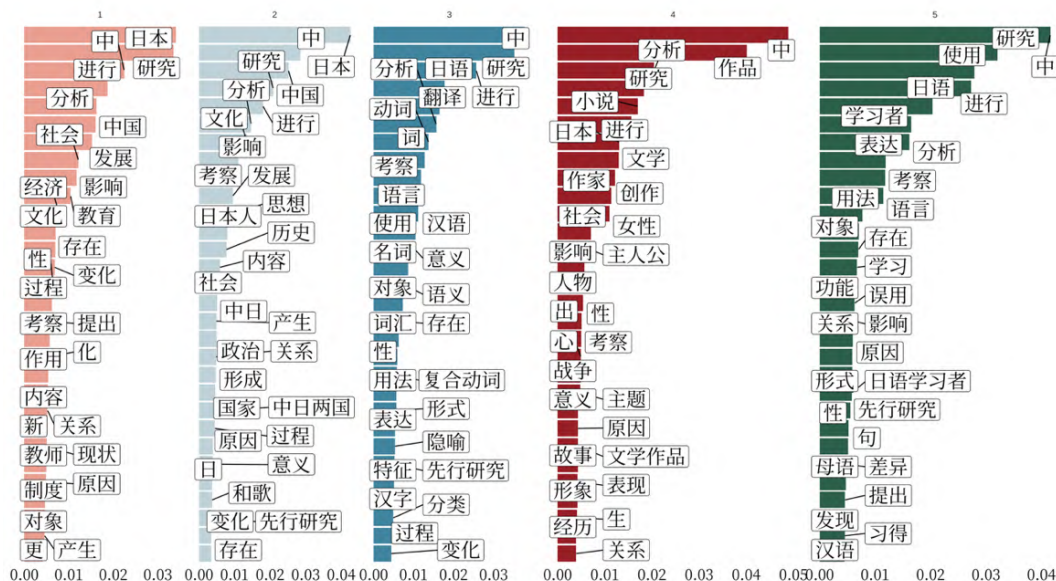


图 8 日语学硕论文五大研究主题

从数据分析可以得知,大多数高校的日语专业硕士论文都以文学研究为主,且文学类研究领域的研究对象重复率较高。语料库技术与语言学之外的研究领域结合不够密切,使用数理化统计思维进行定量研究的硕士论文数量偏少。虽有个别高校在语料库使用上表现突出,但比较集中在语料库与语言学的结合上,语料库技术与文学、国别区域等方面的研究结合较少。通过 LDA 模型分类得到的学科五大研究主题中显示,日语语言文学专业的毕业论文仍然以传统人文研究为主,缺乏对日本社会、国别等领域的关注,此类“非语言文字”类的研究只有在少数硕士论文中出现。

(二)展望

日语语言文学研究不应只以语言文学为主,应该将更广泛、更具涵盖力的日本学、东亚学纳入学科研究的范畴。^[4]但实现此目标不仅需要日语语言文学专业研究生的自我驱动,还需要各大高校建设、发展和对接新文科研究的日语专业研究生的师资队伍,设置除语言课程以外的其他课程,与自然语言处理等领域进行融合、促进传统文科的文文交叉、文理交叉。^[1]此外,应当合理设置日语学科课程,积极促进外语学科与其他学科之间的交流,促进学科教育的改革与长足发展。^[5]在传统文科面临巨大变革的背景下,硕士作为在高校中最具有创新力的群体之一,应该积极强化对不同知识领域的了

解、学习,在继承传统人文研究的基础上,积极从“本体研究”走向“应用研究”,积极使用跨学科方法,结合跨学科知识进行研究。日语语言文学专业研究生在硕士论文选题时应当避免已经重复研究的内容,将视野投向更广阔的跨学科研究方向,例如与心理学、社会学、科技等学科和领域进行交叉,贴合新文科发展的潮流,使得日语学科的研究能够更好地对接国家和社会对日语高等教育人才的需求。此外,日语语言文学硕士还应该积极掌握数字化技术的应用方法,在研究中尝试数字人文视角的定量研究范式。

参考文献:

- [1] 胡开宝.新文科视域下外语学科的建设与发展:理念与路径[J].中国外语,2020,17(3):14-19.
- [2] 闫鑫.中国日语专业学术研究生硕士论文选题过程实证研究[D].北京:北京外国语大学,2020:3-4.
- [3] 朱亮,孟宪学.文献计量法与内容分析法比较研究[J].图书馆工作与研究,2013(6):64-66.
- [4] 王升远,黄彩霞.流弊与突围:“日语语言文学”学科的名与实[J].探索与争鸣,2017(5):83-90.
- [5] 王升远,修刚,王忻,等.中国的日语语言研究:困境、挑战与前景[J].日语学习与研究,2021(5):1-20.

(责任编辑:陈奕颖)