

Doi:10.20063/j.cnki.CN37-1452/C.2023.01.009

基于法律文书语料库的 数词“二”与“两”用法差异分析

孙佳

(中国社会科学院大学 文学院,北京 102488)

摘要:“二”和“两”的用法差异研究一直被学者关注,但以往大多是基于定性的研究,而针对特定领域的研究还不够深入。“二”和“两”在两个语料库的使用存在较大差异,法律语言须准确详细描述事实,多使用数词“二”,体现了法律语言的精准性;法律文书中注重细节描写,常用“数词+量词”这种搭配,且“二”和“两”有相对明确的分工,在选择量词时有固定倾向,而国家语委语料库中,“二”和“两”在选择量词时相对自由,体现了法律语言的稳定性;此外,法律语料库中更倾向于使用省略量词的说法,映射出法律语言的简洁性。

关键词:“二”;“两”;用法差异;法律文书;语料库

中图分类号:H08 **文献标志码:**A **文章编号:**1673-8039(2023)01-0061-07

数词“二”和“两”在意义上都表示数量2,但用法同中有异。“二”和“两”以及两者的差异一直被研究者从多个角度来分析,如古今差异、语境差异等。但专业领域的数词使用研究是否能支持专业领域语言特点的相关研究则较少。

法律文本具有规范性,包括精准性、稳定性和简洁性^[1]。法律语言的精准性要求用词严谨,特别是数词和量词的使用必须严谨,给案件结果提供直接证据,所以对法律文书中数词的研究具有重要意义。加之法律文书的专业性,可能导致“二”和“两”以及两者的使用差异较普通文本有所不同,所以基于法律文书语料库的“二”和“两”以及两者的使用差异研究具有实用价值,对于立法从业者和法律语言研究者在选择数词“二”和“两”时有指导意义,同时有助于法律语言规范化发展。

本文采用定量为主、定性为辅的研究方法,基于语料库研究“二”和“两”在法律文书语料库中的用法差异。4005万字的法律文书语料库由刑法法律文书构成,数据由案情描述和事实部分组成。同时引入3447万字的国家语委语料库,考察这些差异是否同时存在以及这些差异在法律文书领域的特殊性。

我们试图回答以下问题:

其一,“二”和“两”在法律文书语料库中频数、量词搭配、无量词搭配的差异有哪些?为什么?

其二,“二”和“两”在法律文书语料库中频数、量词搭配、无量词搭配的差异是否同时体现在国家语委语料库中?为什么?

其三,较国家语委语料库,法律文书语料库中“二”和“两”差异的特殊性体现了法律语言的什么特点?

一、研究现状

关于数词“二”和“两”以及两者的对比研究有不少。吕叔湘指出“二”和“两”在用法上有明确的分工^[2];王希杰通过举例的方式详细列举出了“二”和“两”的使用语境^[3];朱德熙提到“二”在数数时、系位组合的末项和在“第”之后使用的情况是“两”不能代替的,但在做系数和与量词搭配时,二者均可使用,但“十”之前只能用“二”^[4]⁵⁸;惠兰提到“二”和“两”所指的数目是相同的,但在使用上,二者的用法不尽相同,除了表示数目之外,它们还有各自的含义^[5];李春燕从语言和文化的角度探讨“二”字的词义及其历时

收稿日期:2022-10-12

基金项目:国家语委“十三五”科研规划2020年度重点项目(全球中文学习联盟研究专项)“面向国际中文教育的文本可读性智能评价方法研究及分析系统构建”(ZD1135-141)

作者简介:孙佳(1988—),女,湖北天门人,中国社会科学院大学文学院博士研究生。

变化以及与“两、俩、再”的用法区别^[6]。徐秀花从历时研究的角度比较了“二”和“两”的古今用法^[7];林芮莹通过对数词“2”的古今变化来看文化的演变^[8];李凤吟、张星参考多部词典对“二”和“两”从含义上做了分析对比来看中国的数字文化^[9]。也有更加细化的研究,如张其昀等专门针对“二”和“两”表示不定数,探讨了它们的用法^[10];张素宁对“两”和“二”直接修饰名词或名词性语素的使用情况进行了对比研究^[11]。也有从对外汉语教学角度进行“二”和“两”用法分析的^[12-13]。

但这些研究都是基于定性的研究,基于语料库的定量研究还较少。程锦基于语料库分析了“二”和“两”在新的网络语言时代呈现的新特征,并试图从社会和文化心理角度对这些特征进行解释^[14];蒋妍基于北京语言大学 HSK 语料库,运用历时与共时、形式与意义相结合的方法对“二”类词做宏观的梳理并对其进行微观的分析,归纳出“二”类词在古今汉语和中外语言中语法、语义、语用上的不同,并在此基础上统计出留学生在在使用“二”类词时所出现的偏误,分析偏误产生的原因,最后制定出相应的教学策略^[15]。

针对专业领域内数词的研究相对较少,而“二”和“两”的研究则是少之又少。高一飞以法律数字为研究对象,着重探讨了法律数字的原理、性质、标尺、价值与合理性问题^[16];赵海翔通过结合词典的专业分析和人们的日常用语习惯,建议法律语言中使用“两年”而不是“二年”这种说法^[17];韩桂珍指出法律文书中,数量词对于罪刑的判断有绝对的作用,所以必须准确使用^[18];郭龙生举例说明了在法律语言中“二”和“两”与量词搭配的不同情况^[19]。

以上针对专业领域的数词研究均是基于定性的研究,在专业领域中,特别是针对“二”和“两”的定量研究暂未发现。基于此,我们针对法律文书,结合语料库采用定量为主、定性为辅的研究方法对“二”与“两”的用法差异进行分析。

二、研究所用语料库

法律语言的选择是严谨的,要受制于立法、司法实践及社会价值,所以法律文书语料库有其独特之处。接下来简要介绍法律文书语料库与国家语委语料库的基本信息以及法律文书的语言特点。

(一)法律文书语料库和国家语委语料库

1. 法律文书语料库

其建立主要分为以下几个步骤。

(1)数据来源。法律文书语料库的数据集来自于中国法研杯比赛,是“中国裁判文书网”公开的刑事法律文书,其中每份数据由法律文书中的案情描述和事实部分组成,同时也包括每个案件所涉及的法条、被告人被判的罪名和刑期长短等要素。

(2)数据清洗。去掉语料中的重复数据,对乱码进行处理,将标点符号由半角转为全角等。

(3)分词和词性标注。基于整理完的语料,采用“结巴”中文分词工具对整理完的语料进行分词和词性标注。

2. 国家语委语料库

(1)数据来源。国家语委语料库是一个大型、通用的现代汉语平衡语料库。国家语委语料库由国家语委委托语言文字应用研究所负责建设和管理,于2012年建立。语料分类为人文与社会科学类(政治、历史、社会、经济、文学、艺术等)、自然科学类(农业、工业、医学、电子、工程技术等)和综合类(应用文、难于归类的语料)。语料库的样本分布年份从1919年到2002年。

(2)数据清洗,分词和词性标注。由于国家语委语料库本身经过严格的机助人校的加工方式,已经包括分词和词性标注等信息,所以这里不做额外的处理。

法律文书语料库和国家语委语料库的基本信息如表1所示。

表1 法律文书语料库与国家语委语料库的基本信息

	法律文书语料库	国家语委语料库
总字数	40 052 909	34 774 648
总词数	23 009 224	22 435 006
样本数	83 675	19 923
句子数	647 934	864 320
标点符号	3 717 441	3 812 536

(二)法律文书的语言特点

法律的专业性意味着法律语言有其独有的特点。法律文本具有规范性,包括精准性、稳定性和简洁性^[1]。从语言学的角度,法律的精准性要求立法者在遣词造句方面秉持法律文本语言准确和庄重的格调;稳定性体现在法律文本的用词习惯应在一定时期内保持稳定,不宜随意变动;简洁性体现在法律文本用尽可能少的语言描述全面的

信息。

型例比 TTR (TYPE-TOKEN RATIO) 是衡量文本词汇丰富度的一个指标。TTR 越大表示该文本使用的词汇越丰富。吉罗指数 (Guiraud's index) 在型例比上进行优化,使其不受文本大小的影响,其统计结果如表 2 所示。

表 2 法律文书语料库与国家语委语料库中的吉罗指数

	法律文书语料库	国家语委语料库
吉罗指数	48.72	51.46

可以看出,法律文书语料库中的词汇丰富度弱于国家语委语料库。但是即便如此,法律文书语料库中的数词占比却远高于国家语委语料库中的数词占比,如表 3 所示,这是因为法律文本中多是对事实的陈述、对案件结果的描述,必须要准确详细。

表 3 法律文书语料库与国家语委语料库中的数词占比

	法律文书语料库	国家语委语料库
数词占比	12.42	3.38

三、法律文书语料库中“二”与“两”的用法差异分析

下面将从“二”和“两”的频数、量词搭配和无量词搭配三个方面来定量分析二者之间的差异,同时考察在国家语委语料库和法律文书语料库中“二”和“两”的差异是否有其特点。

(一) “二”和“两”的频数

频数是指词汇在语料库中出现的次数。“二”和“两”在法律文书语料库中的频数信息如表 4 所示。

表 4 “二”和“两”在法律文书语料库中的频数

	法律文书语料库
“二”出现的频数	51,569
“两”出现的频数	23,274

可以看到,“二”出现的频数是“两”的 2.22 倍。在法律文书语料库中“二”的使用较“两”更加常见。

由于法律文书语料库和国家语委语料库的规模略有不同,为了使其拥有可比性,引入标准化频数。标准化频数是指词汇在语料库中出现的占比,将频数归到一个共同的基数。同样,在国家语委语料库中,“二”和“两”的标准化频数也存在不同的分布,“两”的出现更加频繁。为了能够进行更加全面的比较,可以从统计计量学角度,通过对比数据之间的差异是否具有显著性的方法加以综

合检验。使用卡方检验不仅可以考虑词汇在两个语料库中出现的频数以及语料库的大小,还考虑到了词汇在语料库中不出现的情况。最终的统计结果如表 5 所示。

表 5 “二”和“两”的标准化频数和卡方值

	法律文书语料库	国家语委语料库	卡方值
“二”的标准化频数	1.29	1.16	252.43
“两”的标准化频数	0.58	1.22	8687.08

卡方值越大,两者的差异越显著。“二”和“两”在两个语料库中的使用存在差异,并且“两”的差异更加显著。

值得注意的是,《现代汉语词典》(第 7 版)指出“二”有以下特定用法不能用“两”代替:①读数目字只用“二”不用“两”;在小数和分数中用“二”;在多位数中,百、十、个位用“二”不用“两”。②序数也只用“二”^{[20]815}。分别统计这两种情况在语料库中的占比,如表 6 所示。

表 6 “二”在法律文书语料库与国家语委语料库中的使用占比

单位%	法律文书语料库	国家语委语料库
①表数目	24.21	30.12
②表序数	22.76	33.93
③其他	53.03	35.95

可以观察到,“二”在法律文书语料库中,特有用法占比为 46.97%,而在国家语委语料库中占比为 64.05%。

(二) “二”和“两”的量词搭配

量词是表计量的单位,其基本功能是和数词结合起来表示数量。统计分析发现,在法律文书语料库中,“二”的后接量词搭配占其所有后接词中的 30.11%,“两”的后接量词搭配占 76.19%。相较于“二”,“两”更多地与量词搭配。按照频率,表 7 是“二”和“两”最常见的量词。

表 7 法律文书语料库中“二”和“两”的高频量词搭配

	量词 Top10
二	轮、份、个、只、张、组、台、名、袋、盒
两	轮、个、次、部、名、辆、张、包、把、支

表 7 中共出现量词 16 个,其中“轮、个、张、名”这 4 个量词是两者都比较常搭配的,占比 25%。

通过计算“两+量词”搭配与“二+量词”搭配的频率比例,可观察当前量词更倾向于和哪个数词搭配,称为“量词搭配倾向性计算”。量词倾向性取值大于数值 1 表示当前量词更倾向于和

“两”搭配,反之,量词倾向性取值小于数值1表示当前量词更倾向于和“二”搭配。具体如图1所示。

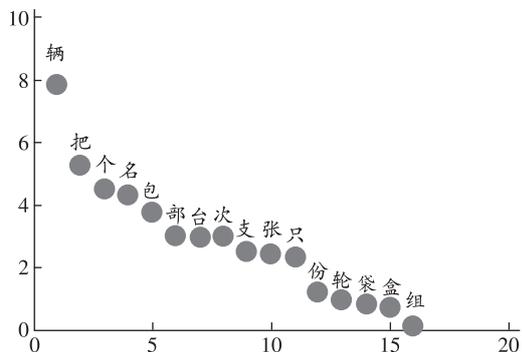


图1 法律文书语料库中量词的搭配倾向性

可以看到,量词“辆”“把”“个”“名”“包”“部”“台”“次”“支”“张”“只”“份”更倾向于和“两”搭配,而“轮”“袋”“盒”“组”更倾向于和“二”搭配。

在国家语委语料库中,“二”的后接量词搭配占其所有后接词中的15.35%,“两”的后接量词搭配占50.77%。按照频率,表8是“二”和“两”最常见的量词。

表8 国家语委语料库中“二”和“两”的高频量词搭配

	量词 Top10
二	个、种、条、元、位、只、袋、颗、根、道
两	个、种、次、只、条、位、块、项、张、颗

表8中共出现量词14个,其中“个”“种”“只”“条”“位”“颗”这6个量词是两者都比较常搭配的,占比30%。在国家语委语料库中,“二”和“两”的搭配倾向性,如图2所示。

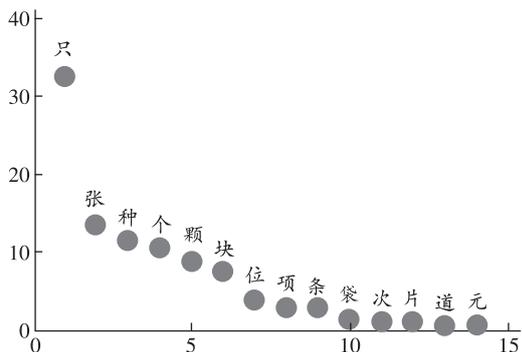


图2 国家语委语料库中量词的搭配倾向性

可以看到,在国家语委语料库中,量词“只”“张”“种”“个”“颗”“块”“位”“项”“条”“袋”“次”“片”更倾向于和“两”搭配,而“道”“元”更倾向于和“二”搭配。

从“二”和“两”与量词搭配的占比可以看到,在国家语委语料库中,相较于“二”,“两”同样更

多与量词搭配,但“二”和“两”与量词搭配情况在两个语料库中也存在差异:①在法律文书语料库中,“二+量词”和“两+量词”的搭配比例均大于国家语委语料库中的比例;②“二”和“两”量词搭配占比的比例有所不同,在法律文书语料库中,“两”的量词搭配占比是“二”的2.53倍,而在国家语委语料库中,其比例扩大至3.30。

从高频量词来看,法律文书语料库会更多出现“数词+量词+物件”这种搭配,如“二轮摩托车、二份证据、二只手镯”;而在国家语委语料库中,“数词+量词”后接的搭配就比较多样,如“二袋水果”“二碗米饭”“二节课”等。

量词搭配倾向性取值越接近数值1,表示在搭配同一量词时,“二”和“两”的使用界限越模糊,两者通用性高;反之,量词搭配倾向性取值越偏离数值1,则表示“二”和“两”的使用界限越清晰,如“二个”和“两个”的比例为1.02,说明“二”和“两”均可以和“个”搭配;而“两辆”和“二辆”的比例为7.87,说明“两”与“辆”搭配更加固定。分别统计在两个语料库中所有量词搭配倾向性取值的平均值可以观察“二”和“两”在搭配同一量词时的使用界限,具体信息如表9所示。

表9 量词搭配倾向性取值的平均值

	法律文书语料库	国家语委语料库
平均值	25.73	12.68

从表9可以观察到,法律语料库中的量词搭配倾向性取值的平均值大于国家语委语料库中量词搭配倾向性取值的平均值。说明在法律文书语料库中,“二”和“两”有相对明确的分工,“二”和“两”在选择量词有固定倾向,“数词+量词”的搭配更加明确。而国家语委语料库中,对于同一量词,既可以与“二”搭配,又可以与“两”搭配,“二”和“两”在选择量词时相对自由。

(三) “二”和“两”无量词搭配

数词除了和量词搭配以外,有时也会省略量词直接与其他词类搭配,如直接修饰名词作为定语,直接修饰动词作状语。在法律文书语料库中“二”和“两”后搭配的词类见图3。

对于“二”,搭配名词占比28.43%,搭配动词仅1.3%;而对于“两”,除了量词之外,最常见的搭配词类也是名词,搭配名词占比19.27%,搭配动词占比0.52%。表10是“二”和“两”在法律文书语料库中的高频名词搭配。

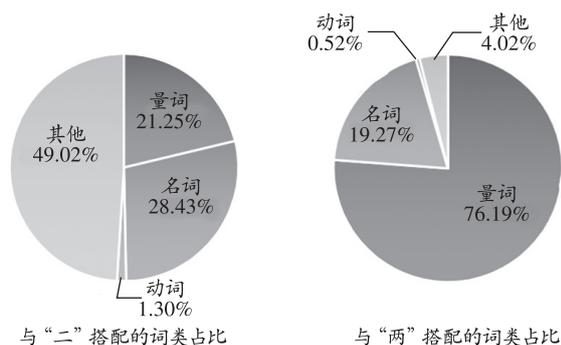


图3 法律文书语料库中“二”和“两”后搭配的词类占比

表10 法律文书语料库中“二”和“两”高频搭配之名词

	名词搭配 Top5
二	人、楼、被告人、被害人、罪
两	人、车、罪、被告人、被害人

可以看到,“二”和“两”均可以与人、物相关的名词搭配。

而在国家语委语料库中,对于“二”,搭配名词占比31.05%,搭配动词5.48%,搭配量词比搭配量词更加常见;而对于“两”,搭配名词占比40.44%,搭配动词占比仅1.25%,除了量词之外,最常见的搭配词类也是名词。详见图4。

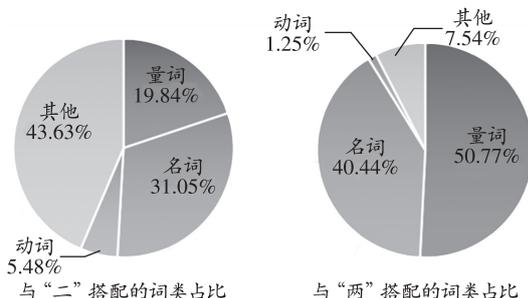


图4 国家语委语料库中“二”和“两”后搭配的词类占比

由于法律文书语料库中词汇丰富度较低些,其中“二+名词”搭配少于国家语委语料库中“二+名词”搭配,同时“两”更多搭配量词,所以其搭配名词能力弱于国家语委语料库中的搭配名词能力。在国家语委语料库中,高频搭配名词如表11所示。

表11 国家语委语料库中“二”和“两”高频搭配之名词

	名词搭配 Top5
二	天、人、倍、字、代
两	年、人、天、方面、类

对比表10和表11不难发现,在法律文书语料库中,“二”和“两”的搭配词限于人、物、伤害动作、地点等,而国家语料语料库中,“二”和“两”的搭配词更加丰富。

“二人、两人”的说法在法律文书语料库和国

家语委语料库均高频出现,具体占比如图5所示。

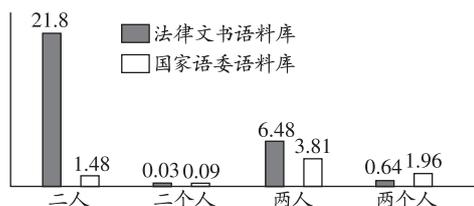


图5 “二人、二个人、两人、两个人”在法律文书语料库和国家语委语料库中的占比

可以看到,“二人、两人”的说法在法律文书语料库中的占比均大于国家语委语料库中的占比。更进一步观察,两个语料库中“数词+人”和“数词+量词+人”的比例可以得出哪个语料库更倾向于使用省略量词的说法。具体统计结果如表12所示。

表12 “二人与二个人、两人与两个人”在法律文书语料库和国家语委语料库中频率占比

单位%	法律文书语料库	国家语委语料库
“二人”与“二个人”频率比例	726.67	16.44
“两人”与“两个人”频率比例	10.12	1.43

通过以上数据发现,法律文书语料库中更倾向于使用省略量词的用法。

四、“二”与“两”用法差异的语言学分析

这一节从“二”和“两”的频数、量词搭配、无量词搭配三个方面来分析二者在法律文书语料库和国家语委语料库中差异的原因,并进一步分析法律文书中这些差异的特殊性体现了法律语言所具有的特点。

(一)“二”和“两”词频差异分析

在法律文书语料库中,“二”比“两”出现的频率高,但在国家语委语料库中,“两”比“二”出现的频率高。因为法律文书会出现“伤势+二级”这样的高频词组,直接导致了“二”在法律文书中出现的频率更高,这体现了法律文书用词精准的特点。通过卡方值显示,“二”的使用频率在两个语料库中差别很大,但“两”的使用频率在两个语料库中差别更大。因为在法律文本中多是对事实的描述和对案件结果的描述,必须准确详细,所以会比较多地用到数词,这会直接导致“二”和“两”在两个语料库的频率出现较大的差异。值得说明的是,“二”较“两”有独有用法,如读数目字、小数、分数以及在多位数中处于百、十、个位时。据

统计,“二”的非特定用法在法律文书中占比高于国家语委语料库。对数词的多样化使用更进一步体现了法律语言的精准性。

(二)“二”和“两”量词搭配差异分析

在法律文书语料库中,相较于“二”,“两”与量词搭配更加灵活,不同的量词有其更具倾向性的数词选择,比如“辆”“把”“个”“名”“包”“部”“台”“次”“支”“张”“只”“份”更倾向于和“两”搭配,而“轮”“袋”“盒”“组”更倾向于与“二”搭配。“两”比“二”与量词搭配更加灵活同样体现在国家语委语料库中,“两”和“二”的量词搭配倾向性为:“只”“张”“种”“个”“颗”“块”“位”“项”“条”“袋”“次”“片”更倾向于和“两”搭配;而“道”“元”更倾向于和“二”搭配。但法律文书语料库中“两+量词”和“二+量词”的占比均大于国家语委语料库,因为在法律文书中会注重细节描写,会更多地使用“数词+量词”这种搭配,这体现了法律语言用词严谨的风格。法律文书语料库中“两+量词”与“二+量词”的比例小于国家语委语料库中的比例。因为“二+量词”搭配在国家语委语料库中减少的比率高于“两+量词”搭配在国家语委语料库中减少的比率。

从高频量词来看,法律文书语料库会更多出现“数词+量词+物件”这种搭配中的量词,比如说“二轮摩托车”“二份证据”“二只手镯”;在国家语委语料库中,“数词+量词”后接的搭配就比较多样,如“二袋水果”“二碗米饭”“二节课”等。这与法律文书语料库中的语料注重事实描述有关,也与国家语委语料库的通用性有关。

法律文书语料库中的量词搭配倾向性取值的平均值大于国家语委语料库中量词搭配倾向性取值的平均值,这说明在法律文书语料库中,“二”和“两”有相对明确的分工,“二”和“两”在选择量词方面有固定倾向,“数词+量词”的搭配更加明确。而国家语委语料库中,对于同一量词,既可以与“二”搭配,又可以与“两”搭配,“二”和“两”选择量词相对自由,这符合法律文本的稳定性的特点,体现了法律文本的用词习惯在一定时期内保持稳定。

(三)“二”和“两”无量词搭配差异分析

在法律文书语料库中,“二”更多地搭配名词,其次是量词;而“两”主要是搭配量词,其次是名词。由于法律语料来自刑事案件的案情描述和事实部分,“两+伤害动作”这样的搭配比较常见。

同时“数词+人物”这样的搭配也比较多,比如“二受害人”“两被害人”等,在这种搭配里面,“二”和“两”可以通用。在国家语委语料库中,“二”和“两”搭配的名词和量词占比有相同的规律。但“二”和“两”的搭配词不再局限于人、物、伤害动作、地点等,搭配的名词种类分布更加广泛。

通过对“二人”“二个人”“两人”“两个人”在法律文书语料库和国家语委语料库中的占比分析,可以看到,“二人”“两人”的说法在法律文书语料库中的占比均大于国家语委语料库中的占比。法律文书语料库中“二人”与“二个人”的比例和“两人”与“两个人”的比例均大于国家语委语料库中的比例。这反映了法律文书语料库中更倾向于使用省略量词的说法。这一点正好体现出法律语言的简洁性,用尽可能少的文字表达准确的内容。

由于法律文书语料库采用纯机器标注,而且标注的标准和国家语委语料库的标注标准存在些许差异,这可能会导致分析结果存在微量偏差。因而本研究将在后续工作中将对法律文本的标注质量进行人工校验。同时会尝试从更多的维度来比较“二”和“两”的用法差异,比如与数词的搭配等。此外,可将“二”和“两”与量词的搭配进行更深入的研究。

参考文献:

- [1] 罗德刚. 论法律文本的特点[J]. 知识经济, 2011(8).
- [2] 吕叔湘. “二”和“两”[J]. 语文世界, 1998(11).
- [3] 王希杰. 说“二”道“两”[J]. 汉语学习, 1982(4).
- [4] 朱德熙. 语法讲义[M]. 北京: 商务印书馆, 1982.
- [5] 惠兰. 谈谈“二”与“两”在使用上的区别[J]. 逻辑与语言学习, 1984(1).
- [6] 李春燕. “二”字文化语言学考辨[J]. 现代语文(学术综合版), 2013(12).
- [7] 徐秀花. 数词“二”和“两”古今用法之比较[J]. 山西大同大学学报(社会科学版), 2010(3).
- [8] 林芮莹. 从数词“2”的古今变化看文化演变[J]. 文学教育(上), 2013(8).
- [9] 李凤吟, 张星. 从“二”和“两”看中国数文化[J]. 泉州师范学院学报, 2019(1).
- [10] 张其昀, 谢俊涛. 二、两、俩和再[J]. 南阳师范学院学报, 2008(8).
- [11] 张素宁. “两+N”结构与“俩+N”(“N+俩”)结构

的比较研究[J]. 现代语文(学术综合版),2006(12).

[12] 郭丹凤.“二”和“两”的偏误分析与对外汉语教学研究[D]. 南昌:南昌大学,2016.

[13] 朱嫣红.“二”与“两”的用法及其教学[J]. 文学教育(中),2014(1).

[14] 程锦.基于语料库的汉语数词“二”“两”用法的历时研究[J]. 内蒙古农业大学学报(社会科学版),2010(6).

[15] 蒋妍.“二”类词对比研究与偏误分析[D]. 长沙:湖南师范大学,2012.

[16] 高一飞.现代性视域下的法律数字研究[D]. 武汉:武汉大学,2018.

[17] 赵海翔.“二年”宜改为“两年”[J]. 法学,1993(7).

[18] 韩桂珍.浅谈司法文书的准确性[J]. 河北法学,1994(5).

[19] 郭龙生.浅议立法语言规范[J]. 语言与法律研究,2020(1).

[20] 中国社会科学院语言研究所词典编辑室.现代汉语词典(第7版)[M]. 北京:商务印书馆,2016.

Analysis of Usage Differences of Numerals “Er” (二) and “Liang” (两) on the Basis of Legal Instrument Corpus

SUN Jia

(School of Literature, University of Chinese Academy of Social Sciences, Beijing 102488, China)

Abstract: The studies of usage differences between “er” (二) and “liang” (两) have been concerned by scholars, but most of them are based on the qualitative method, and the study aimed at specific areas is not deep enough. There exist great differences of the two characters' usage in the two corpora. Legal language must describe the facts accurately and elaborately, and more numerals are used, reflecting the accuracy of legal language. Legal instruments pay attention to detailed description and often use the “numeral + quantifier” collocation, there is a relatively clear division of labor between “er” (二) and “liang” (两), and there is a fixed tendency to choose quantifiers. But in the corpus of National Language Commission, “er” (二) and “liang” (两) are relatively free when choosing quantifiers, reflecting the stability of legal language. Besides, the legal corpus tends to use the statement of ellipsis of quantifiers, which reflects the simplicity of legal language.

Key words: “er” (二); “liang” (两); usage difference; legal instrument; corpus

(责任编辑 雪 箫)