

基于语言学中文学术图书的 半衰期分析研究*

程 路

(西华大学 四川学术成果分析与应用研究中心, 四川 成都 610039)

摘要: 文献半衰期能够从文献利用的角度上, 定量地揭示科学文献的老化规律。本文利用语言学学术图书的参考文献和引证数据, 测量中文语言学图书的引用图书半衰期、引用期刊半衰期、被图书引用半衰期、被期刊引用半衰期及其变化趋势, 多角度揭示了图书的老化特征及其演变规律, 对探索文献老化规律、提升图书利用价值、优化馆藏结构具有重要价值。

关键词: 半衰期; 语言学; 中文学术图书

中图分类号: H0; G250.252

文献标志码: A

文章编号: 1004-342(2019)05-79-07

1958年贝尔纳(J. D. Bernal)第一次提出用“半衰期”来表征文献情报老化速度。1960年, 巴尔顿(R. E. Burton)和开普勒(R. W. Kebler)提出了文献“半衰期”的概念, 指现时尚在利用的文献中较新的一半是在多长一段时间内发表的, 也被称为“中值引文年限”。原美国科学信息情报所根据这种定义, 从引证与被引证的角度出发, 又提出了引用半衰期和被引半衰期。文献半衰期能够从文献利用的角度上, 定量地揭示科学文献的老化规律, 衡量科学文献的老化速度和程度。^[1]

经过文献调查发现, 目前关于文献半衰期的研究大多基于期刊文献、学位论文或者标准。方红玲^[2]、杨彭^[3]、卓可秋^[3]、黄利平^[4]等分别基于期刊文献计算了科学、数学和环境科学、艺术学、管理学、图书情报学等学科的半衰期。N. Zafrunnisha和V. P. Reddy^[5]计算了印度的安得拉大学等3所大学哲

学学科博士论文的引用半衰期。A. M. S. Angammana和L. A. Jayatissa^[6]计算了科伦坡大学、凯拉尼亚大学图书情报学科的硕士论文引用论文半衰期。陈云鹏^[7]等人计算了标准文献的半衰期。随着BKCI的发布, 也有学者利用BKCI的数据研究某个学科的半衰期。刘晓娟、马梁^[9]、宋雯斐、刘晓娟^[10]基于BKCI分析了图书的半衰期。但少有基于中文图书研究图书引用半衰期或者被引半衰期, 更少有研究者基于中文图书同时对图书的引用半衰期和期刊半衰期进行研究。

因此, 本文将利用中文语言学图书参考文献数据和引文文献数据研究中文语言学图书的被图书引用半衰期、被期刊引用半衰期以及引用图书半衰期和引用期刊半衰期, 这对补全文献半衰期研究的版图, 探索中文图书的老化特征及其演变规律, 促进图书馆藏、剔旧等由经验管理向科学管理的转变, 评价

收稿日期: 2019-05-11

* 基金项目: 四川省社科规划青年项目“基于国家社科基金项目的四川省哲学社会科学影响力分析研究”(项目编号: SC18C040)。

作者简介: 程 路(1989-), 男, 西华大学四川学术成果分析与应用研究中心讲师。

图书的研究价值,研究图书著者的学术影响力具有重要价值。

一、数据来源

本文首先从国家图书馆获取所有中图分类号以“H”开头的图书,然后根据中图分类号(CLC),过滤非语言学图书;其次按照图书记录中的文献类型字段,过滤所有非“专著”图书;再次按照学术著作标准,过滤教材、手册、科普读物等所有非学术类语言学图书;共获得有效中文语言学图书20 411本。最

后从超星发现数据库获取以上语言学学术著作的参考文献和引证文献。

本文从20 411本图书中选择出版年在1997 - 2014年之间语言学学术图书共16 217本进行半衰期分析。其图书类型参考文献数量、期刊类型参考文献数量分别为3 384 447条、743 793条,在统计年被图书引用总次数、被期刊引用总次数分别为439 766次、370 869次,具体如表1。

表1 1997 - 2014年中文语言学图书引用及其被引分布

年份	数量	图书类型参考文献数量	期刊类型参考文献数量	被图书引用次数	被期刊引用次数
2014	1 344	211 922	30 051	21 285	22 937
2013	1 206	326 523	55 134	44 738	25 415
2012	1 125	234 862	47 046	36 946	30 781
2011	1 140	260 987	56 620	35 289	42 429
2010	1 102	205 500	50 838	31 549	42 959
2009	1 130	322 800	82 502	42 334	40 755
2008	1 051	315 946	79 216	44 411	34 798
2007	896	239 100	60 820	34 657	33 514
2006	954	237 241	57 707	30 052	25 765
2005	897	164 767	38 680	22 046	17 019
2004	842	166 307	35 398	20 137	14 212
2003	791	141 280	30 405	14 495	12 425
2002	840	162 393	34 623	15 499	9 756
2001	714	99 648	20 216	12 475	6 590
2000	642	103 279	21 301	10 474	4 617
1999	537	73 348	16 787	10 185	2 660
1998	508	52 375	10 888	7 756	2 184
1997	498	66 169	15 561	5 438	2 053
合计	16 217	3 384 447	743 793	439 766	370 869

二、半衰期计算

(一) 引用半衰期

文献引用半衰期是指该文献集合在统计年(T年度)中所引用(参考)的文献中,较新文献的引用(参考)次数达到总引用(参考)的50%的年数。文献引用半衰期表示文献参考利用文献的新旧程度。

根据其定义得出对应的计算公式如公式(1)所示:

$$H_{CitingT} = N + (50\% - R1)/(R2 - R1) \quad (1)$$

公式(1)中,文献引用半衰期指数为 H_{Citing} , $R1$ 为累计引用频次最接近、并小于50%的那年的累计引用百分比, $R2$ 为累计引用频次最接近、并不小于50%的那年的累计引用百分比, N 为累计引用百分

比为 R1 的那年到 T 年的间隔年数。根据引用的文献类型的不同,文献引用半衰期可以分为引用期刊半衰期和引用图书半衰期。

引用图书半衰期是指该文献集合在统计年(T 年度)中所引用(参考)的图书中,较新出版图书的引用(参考)次数达到图书总引用(参考)的 50% 的年数。图书引用半衰期表示图书利用图书的新旧程度,其值越小,表示引用的图书越新。其计算方法同公式(1)。

对引用年在 T 年的图书集合进行统计(即统计年为 T),首先把 T 年引用的图书按照出版年度统计,然后分别计算其所占 T 年总引用百分比和累计

百分比,最后将相应 T 年的间隔年数标记为 1,并依次标记间隔年数,分别统计出 1997 - 2014 年图书引用图书出版年分布情况。

根据定义,如表 2 可知,当统计年为 2014 年时,累计引用频次最接近、并小于 50% 的那年的累计引用百分比 $R1 = 49.32\%$,累计引用频次最接近、并不小于 50% 的那年的累计引用百分比 $R2 = 52.88\%$,累计引用百分比为 R1 的那年到 T 年的间隔年数 $N = 12$,根据公式(1),图书在 2014 年的引用图书半衰期 $H_{Citing2014} = 12 + 50\% - 49.32\%)/(52.88\% - 49.32\%) = 12.19$ 。

表 2 引用年为 2014 年语言学图书引用图书半衰期

出版年	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002
文献条数	0	8 180	6 163	6 407	10 355	13 031	9 924	11 125	11 194	8 724	11 332	8 076	7 548
所占比	0	3.86	2.91	3.02	4.89	6.15	4.68	5.25	5.28	4.12	5.35	3.81	3.56
累计百分比	0	3.86	6.77	9.79	14.68	20.83	25.51	30.76	36.04	40.16	45.5	49.32	52.88
间隔年数	1	2	3	4	5	6	7	8	9	10	11	12	13

同理可计算出来源图书在 1997 - 2013 年的引用图书半衰期,计算结果如表 3 所示。

表 3 语言学图书引用图书半衰期

引用年	引用图书半衰期	引用年	引用图书半衰期
2014	12.19	2005	12.14
2013	11.64	2004	12.5
2012	11.04	2003	13.47
2011	10.74	2002	16.72
2010	10.49	2001	14.8
2009	10.65	2000	17.65
2008	11.48	1999	19.14
2007	12.04	1998	19
2006	11.99	1997	18.41

由表 3 可知,在 1990 年 - 2014 年之间,语言学图书最大引用图书半衰期出现在 1999 年,是 19.14 年,最小的出现在 2010 年,是 10.49 年,平均引用图书半衰期是 13.67 年。

引用期刊半衰期是指该文献集合在统计年(T 年)中所引用(参考)的期刊论文中,较新发表期刊论文的引用(参考)次数达到期刊总引用(参考)的 50% 的年数。引用期刊半衰期表示图书利用期刊的新旧程度,其值越小,表示利用的期刊越新,其计算方法同公式(1)。

根据定义,由表 4 可知,当统计年为 2014 年时,累计引用频次最接近、并小于 50% 的那年的累计引用百分比 $R1 = 49.76\%$,累计引用频次最接近、并不小于 50% 的那年的累计引用百分比 $R2 = 54.08\%$,累计引用百分比为 A 的那年到 T 年的间隔年数 $N = 10$ 。根据公式(1),图书在 2014 年的引用期刊半衰期 $H_{Citing2014} = 10 + 50\% - 49.76\%)/(54.08\% - 49.76\%) = 10.06$ 。

同理,可计算出中文语言学图书在 1997 - 2013 年的引用期刊半衰期,计算结果如表 5 所示。

由表 5 可知,在 1997 年 - 2014 年之间,语言学

图书最大引用期刊半衰期出现在 2002 年,是 17.49 年;最小的出现在 2009 年,是 7.46 年;平均引用期刊半衰期是 11.94 年。

表 4 引用年为 2014 年语言学图书引用期刊半衰期

出版年	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004
文献条数	1 352	1 780	1 710	1 675	1 354	1 602	1 401	1 593	1 261	1 224	1 300
所占比	4.50	5.92	5.69	5.57	4.51	5.33	4.66	5.30	4.20	4.07	4.33
累计百分比	4.50	10.42	16.11	21.69	26.19	31.52	36.19	41.49	45.68	49.76	54.08
间隔年数	1	2	3	4	5	6	7	8	9	10	11

表 5 语言学图书在 1997 - 2014 年引用期刊半衰期

引用年	引用期刊半衰期	引用年	引用期刊半衰期
2014	10.06	2005	11.21
2013	9.64	2004	12.83
2012	9	2003	15.08
2011	8.32	2002	17.49
2010	8.08	2001	15.5
2009	7.46	2000	15.51
2008	8.35	1999	17.05
2007	8.65	1998	15.57
2006	10.06	1997	15.21

(二)被引半衰期

文献被引半衰期,指文献集合在统计年(T 年度)被引用的总次数中,较新的一半文献是在多长时间内发表的,这个时间段就是文献的被引半衰期。文献被引半衰期越长,说明文献的学术生命也越长,该文献越有长期的参考价值。根据其定义得出对应的计算公式如公式(2)所示:

$$H_{\text{CitedT}} = M + (50\% - C1)/(C2 - C1) \quad (2)$$

表 6 被引年为 2014 年语言学图书被图书引用半衰期

出版年	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004
文献条数	4	1 718	1 885	805	646	410	702	2 653	764	908	629
所占比	0.02	8.07	8.86	3.78	3.04	1.93	3.30	12.46	3.59	4.27	2.96
累计百分比	0.02	8.09	16.95	20.73	23.76	25.69	28.99	41.45	45.04	49.31	52.26
间隔年数	1	2	3	4	5	6	7	8	9	10	11

根据定义,由表 6 可知,当统计年为 2014 年时,累计被引频次最接近、并小于 50% 的那年的累计引

公式(2)中,文献被引半衰期指数为 H_{Cited} , C1 为累计被引频次最接近、并小于 50% 的那年的累计被引百分比, C2 为累计被引频次最接近、且不小于 50% 的那年的累计被引百分比, M 为累计被引百分比为 C1 的那年到 T 年的间隔年数。根据被引对象的不同类型文献,可以分为被图书引用半衰期、被期刊引用半衰期。

被图书引用半衰期是指该文献集合在统计年(T 年度)被图书引用的总次数中,较新的占 50% 的被引图书的出版间隔年数。被图书引用半衰期表示文献生命力,其值越长,表示其对图书的价值越大,其计算方法同公式(2)。

对被引年在 T 年的图书集合进行统计(即统计年为 T),首先把在 T 年被图书引用的图书按照出版年度统计,然后分别计算其所占 T 年总被图书引用百分比,和累计百分比,最后将相应 T 年的间隔年数标记为 1,并依次标记间隔年数,分别统计出 1997 - 2014 年图书被图书引用出版年分布情况。

用百分比 $C1 = 49.31\%$, 累计被引频次最接近、并不小于 50% 的那年的被引累计百分比 $C2 = 52.26\%$,

被引累计百分比为 C1 的那年到 T 年的间隔年数 $X = 10$, 根据公式(1), 图书在 2014 年的被图书引用半衰期 $H_{Cited2014} = 10 + 50\% - 49.31\%)/(52.26\% - 49.31\%) = 10.23$ 。

同理可计算出来源图书在 1997 - 2013 年被图书引用半衰期, 计算结果如表 7 所示。

由表 7 可知, 在 1997 - 2014 年之间, 语言学图书的最大被图书引用半衰期出现在 2010 年, 是 10.96 年, 最小出现在 1997 年, 是 4.36 年, 平均被图书引用半衰期是 8.53 年。

被期刊引用半衰期是指该文献集合在统计年 (T 年度) 被期刊论文引用的总次数中, 较新的占 50% 的被引图书的出版间隔年数。被期刊引半衰期

表示文献生命力, 其值越长, 表示其对期刊论文的价值越大, 其计算方法同公式(2)。

表 7 语言学图书在 1997 - 2014 年被图书引用半衰期

被引年	被图书引用半衰期	被引年	被图书引用半衰期
2014	10.23	2005	9.58
2013	10.11	2004	8.66
2012	10.44	2003	8.24
2011	10.75	2002	7.64
2010	10.96	2001	6.60
2009	9.91	2000	6.18
2008	9.54	1999	5.59
2007	10.13	1998	4.93
2006	9.76	1997	4.36

表 8 被引年为 2014 年语言学图书被期刊引用半衰期

出版年	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003
文献条数	198	232	611	1 082	726	1 000	1 046	802	1 194	1 666	1 992	1 173
所占比	0.86	1.01	2.66	4.72	3.17	4.36	4.56	3.50	5.21	7.26	8.68	5.11
累计百分比	0.86	1.87	4.54	9.26	12.42	16.78	21.34	24.84	30.04	37.31	45.99	51.11
间隔年数	1	2	3	4	5	6	7	8	9	10	11	12

根据定义, 由表 8 可知, 当统计年为 2013 年时, 累计被引频次最接近、并小于 50% 的那年的累计引用百分比 $C1 = 45.99\%$, 累计被引频次最接近、并不小于 50% 的那年的被引累计百分比 $C2 = 51.11\%$, 被引累计百分比为 C1 的那年到 T 年的间隔年数 $M = 11$, 根据公式(1), 图书在 2014 年的被期刊引用半衰期 $H_{Cited2014} = 11 + 50\% - 45.99\%)/(51.11\% - 45.99\%) = 11.78$ 。

同理可计算出来源图书在 1997 - 2013 年被期刊引用半衰期, 计算结果如表 9 所示。

由表 9 可知, 在 1997 年 - 2014 年之间, 语言学图书最大被期刊引用半衰期出现在 2013 年, 是 11.84 年; 最小的出现在 1997 年, 是 4.31 年; 平均被期刊引用半衰期是 7.28 年。

表 9 语言学图书在 1997 - 2014 年被期刊引用半衰期

被引年份	被期刊引用半衰期	被引年份	被期刊引用半衰期
2014	11.78	2005	6.29
2013	11.84	2004	5.91
2012	10.67	2003	5.98
2011	9.54	2002	5.49
2010	9.03	2001	5.05
2009	8.44	2000	5.00
2008	7.88	1999	4.83
2007	7.18	1998	5.11
2006	6.72	1997	4.31

三、半衰期分析

(一) 引用半衰期分析

1997—2014 年, 中文语言学图书的引用图书半衰期在 10.49 ~ 19.14 年之间, 引用期刊半衰期在

7.46~17.49之间,平均引用图书半衰期13.67年大于平均引用期刊半衰期11.94年,其年度分布曲线如图1所示。

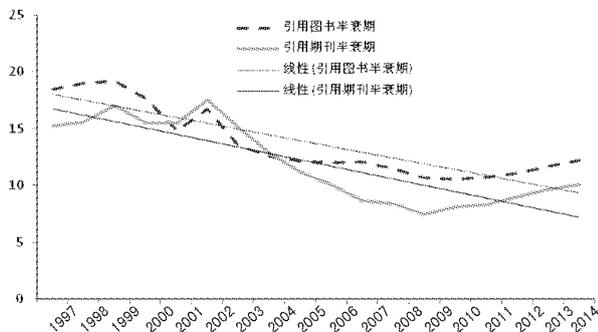


图1 语言学图书引用图书和期刊半衰期

由图1可知,引用图书半衰期和引用期刊半衰期走势相似,进一步分析易知,两者的相关 Pearson 相关系数为 0.90,表明图书的引用图书半衰期和引用期刊半衰期正相关。这给我们测量图书半衰期提供新思路,在现有图书引文数据库不够完善的条件下,当引用期刊文献的引用图书半衰期和引用期刊半衰期正相关时,可以利用期刊引用图书的数据来表征图书的半衰期变化。

Sangam^[11]认为文献数量增长越快,文献老化速度越快、半衰期越短。由前表1可知,语言学学术图书的数量越来越多。在1997-2014年之间,引用期刊半衰期和引用图书半衰期的线性趋势如图1直线所示,图书的引用半衰期在变短,图书老化速度在加快,与 Sangam 的结论基本一致,文献老化与文献的数量有关。

进一步分析可知,引用图书半衰期和引用期刊半衰期在2002年之后,其值大小稳步减小,直至2009年之后小幅上升。表明在2002—2009年,学者利用的文献越来越新,可能的原因是技术影响着学术传播和交流,即随着信息技术的发展,学者获取新文献的方式越来越便捷,学者可以更快地获取新文献,并利用新文献。值得注意的是,在2002—2009年之间,引用期刊半衰期比引用图书半衰期下降得更快,反映学者更倾向于或者更快地利用相对较

新的期刊论文。但在2009年以后,文献的半衰期变高,这是否表明现在发表的文献的价值逐渐减低,抑或信息技术的发展使得学者获取老文献越来越方便,值得进一步研究。

(二)被引半衰期分析

1997—2014年,中文语言学图书的被图书引用半衰期在10.96~4.36年之间,被期刊引用半衰期在4.31~11.84年之间,平均被图书引用半衰期8.53年大于被期刊引用半衰期7.28年,其年度分布曲线如图2所示。

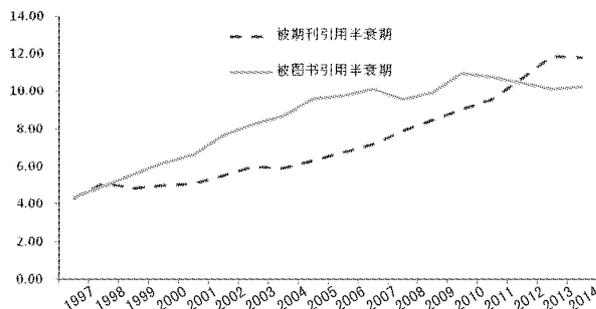


图2 语言学图书被图书和期刊引用半衰期

由图2可知,中文语言学图书的被图书引用半衰期和被期刊引用半衰期走势相似,进一步分析易知,两者的相关 Pearson 相关系数为 0.79,表明图书的被图书引用半衰期和被期刊引用半衰期正相关。

党亚茹等^[12]认为期刊的被引半衰期可以反映研究成果的生命力。由图2可知,中文语言学图书的被图书引用半衰期和被期刊引用半衰期曲线都呈上升趋势,表明语言学学科图书对图书和期刊的利用价值越来越大,生命力越来越强。但是,被图书引用半衰期到2010年开始呈下降趋势,被期刊引用半衰期直至2013年才开始呈微弱下降。

如图2,在1999-2011年之间,被图书引用半衰期大于被期刊引用半衰期,在2012年以后语言学图书的被图书引用半衰期开始小于被期刊引用半衰期。

通过对比分析引用图书半衰期、引用期刊半衰期和被图书引用半衰期、被期刊引用半衰期,本文发现,引用半衰期和被引半衰期表征的半衰期的特征不一样,存在明显区别。

四、结论

本文通过对中文语言学图书的半衰期的实证分析,弥补了中文图书半衰期研究领域的不足。通过测量 1997—2014 年中文语言学图书的引用图书半衰期、引用期刊半衰期、被图书引用半衰期、被期刊引用半衰期,发现引用图书半衰期和引用期刊半衰期、被图书引用半衰期和被期刊引用半衰期分别呈正相关;平均引用图书半衰期 13.67 年大于平均引用期刊半衰期 11.94 年,平均被图书引用半衰期 8.53 年大于被期刊引用半衰期 7.28 年。同时对它们的趋势进行分析,了解到引用半衰期在整体呈下降趋势,但近几年呈上升趋势;被引半衰期整体呈上升趋势,近两年趋于稳定。通过对比图书引文半衰期和被引半衰期,表征的半衰期的特征不一样,存在明显区别。以上结论,对于研究文献的老化规律具有重要价值,为提升图书利用价值、优化馆藏具有重要价值。

由于引用图书的粒度不统一或者规范、图书参考文献著录的不规范以及超星公司提供的参考和引证数据的不全面,致使本文测量的半衰期存在误差,值得进一步采集更全面、准确的数据来对半衰期进行测量。

参考文献:

- [1] 邱均平. 信息计量学[M]. 武汉: 武汉大学出版社, 2007: 79-86.
- [2] 方红玲. 国内外眼科学、数学和环境科学期刊被引半衰期的比较[J]. 中国科技期刊研究, 2018, 29(2): 165-170.
- [3] 杨彭. 基于文献计量学的艺术学学科被引半衰期分析[J]. 西南民族大学学报(人文社会科学版), 2014, 35(11): 228-232.
- [4] 卓可秋. 基于文献计量学的管理学学科半衰期研究[J]. 图书与情报, 2014(4): 55-60.
- [5] 黄利平. 基于引文分析的文献老化研究——以图书情报和管理学为例[J]. 情报杂志, 2011, 30(10): 30-35.
- [6] Zafrunnisha N, Rrddyv P. Citations in Psychology PhD Theses: An Obsolescence Study[J]. Library Philosophy and Practice, 2010(7): 1-8.
- [7] Angammana A M S, Jayatissa L A. A Bibliometric Study of Postgraduate Thesis in Library and Information Science: With Special Reference to University of Kelaniya and University of Colombo, Sri Lanka[J]. Journal of the University Librarians' Association of Sri Lanka, 2015, 19(1): 32-53.
- [8] 陈云鹏, 张宝林. 标准文献半衰期算法及实证研究[J]. 情报杂志, 2012, 31(8): 94-97.
- [9] 刘晓娟, 马梁. 基于 BKCI 的学术著作引文分布研究[J]. 图书情报工作, 2017, 61(24): 105-113.
- [10] 宋雯雯, 刘晓娟. 基于 BKCI 的图书半衰期分析——以图书情报学学科为例[J]. 图书情报工作, 2016, 60(12): 124-129.
- [11] Sangam S. Obsolescence of Literature in the Field of Psychology[J]. Scientometrics, 1999, 44(1): 33-46.
- [12] 党亚茹, 王莉亚. JCR 自然科学版期刊半衰期指标的区间变化分析[J]. 情报科学, 2007, 25(6): 804-810.

(责任编辑: 刘晓红)