

基于语料库的金庸与古龙小说语言风格对比研究

潘纪龙

南开大学汉语言文化学院, 天津

收稿日期: 2022年11月10日; 录用日期: 2022年12月5日; 发布日期: 2022年12月13日

摘要

金庸与古龙为中国文学贡献了大量优质且风格迥异的篇章, 二者魅力无限的语言特点具有分析论证与对比研究的重大价值。本文通过建立武侠小说文本语料库, 利用定性分析与定量分析相结合的计量统计方法, 进行语言风格的细致探究与对比分析, 探究二者语言风格的异同。最终, 我们发现金庸文本受传统文化的浸润更深厚, 体现昂扬的斗争姿态和浓郁的家国情怀, 而古龙则长于对侠客个体形象的书写与描摹, 侧重营造孤高傲然的侠客形象及清冷寂寞的叙事环境。

关键词

语料库, 金庸, 古龙, 语言风格, 高频词

A Corpus Based Comparative Study of the Language Styles of Jin Yong's and Gu Long's Novels

Jilong Pan

College of Chinese Language and Culture, Nankai University, Tianjin

Received: Nov. 10th, 2022; accepted: Dec. 5th, 2022; published: Dec. 13th, 2022

Abstract

Jin Yong and Gu Long have contributed a lot of high-quality and different styles to Chinese literature. Their charming language features are of great value for analysis, demonstration and comparative study. This paper explores the similarities and differences between the two languages

through the establishment of a text corpus of martial arts novels, and the careful exploration and comparative analysis of language styles by using the quantitative and qualitative analysis methods. Finally, we find that Jin Yong's text is more deeply infiltrated by traditional culture, reflecting a high attitude of struggle and a strong sense of family and country, while Gu Long is good at writing and describing the individual image of a chivalrous man, focusing on creating an image of a chivalrous man who is aloof and proud and a cold and lonely narrative environment.

Keywords

Corpus, Jin Yong, Gu Long, Language Style, High-Frequency Words

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

汉语的江湖从不寂寞。20 世纪 50 年代,以金庸和古龙为代表的新派武侠小说风靡华人世界。金庸、古龙笔下丰富的优质作品造成了中国武侠小说历史上独特的“双子星”现象,武侠小说的崛起也引发了海内外学者的高度关注与广泛讨论。

早在 20 世纪末,徐岱(1998) [1]便从艺术价值角度指出金庸武侠小说中“那种史诗般的气概与理想主义品格之难能可贵”。21 世纪以来,芦海英(2006) [2]从人物形象角度出发,认为古龙“通过浪子形象的塑造完成了对人性的挖掘和诠释”。王开银(2008) [3]则通过举例论证,从句段及篇章角度对二者作品进行对比分析。胡菁娜(2013) [4]从题材、语言、情怀三方面进行二者的风格比较。赵铭(2016) [5]则立足于社会与文化因素,分析二者文本异同的原因所在。

然而,此前对二位作家作品的分析研究多集中于历史文化、社会背景等文学艺术角度。从语言学角度,尤其是利用计量分析与统计手段进行的对比研究仍显分量不足。由此,我们产生了试图改善此领域研究空白现状的想法。

在本文中,我们通过使用数理统计工具,计划将从常用句式、高频词、色彩词等角度研究两位作家的语言特点与用语风格。希望借此帮助读者纠正之前模糊的、笼统的感性认识,建立对两位作家的作品之系统性、理性化的认识,以便更加直观地感受两种语言风格的不同魅力。

2. 前期准备

2.1. 文本选取与语料库建立

在当代文学史上,金庸与古龙均属于“高产”作家。据统计,金庸创作武侠小说十五部,古龙则撰写武侠小说共计七十多部。一方面,为了精简统计结果,我们计划选取最受大众欢迎的代表性作品。另一方面,为了使语料库的统计结果能从整体上反映作者的语言风格特点,本次选取二者武侠小说的基本原则是:在选取经典文本的同时,尽量囊括两位作家前、中、后期不同时间段的长、中、短篇作品。

基于以上原则,本次选择了金庸与古龙各自的 5 部代表作,分别为:金庸《射雕英雄传》、《白马啸西风》、《倚天屠龙记》、《鹿鼎记》、《越女剑》;古龙《苍穹神剑》、《飘香剑雨》、《绝代双骄》、《陆小凤传奇》系列(以下简称《陆小凤传奇》)、《风铃中的刀声》。

通过在网络上搜索对比,我们选择了较完整的、错漏较少的版本,并下载相应的文本文件。之后,我们对其中明显的编排错误或上传过程中的乱码现象、格式错误进行人工纠正,力求建立能真实反映作者语言风格的语料库。

最终,我们建立了金庸、古龙武侠小说代表作语料库,并得到关于十部作品总字数、总词数的统计表,如表 1 所示。

Table 1. Statistics of the number of words and phrases in Jin Yong's and Gu Long's works
表 1. 金庸、古龙各自作品的字数、词数统计表

金庸作品	总字数	总词数	古龙作品	总字数	总词数
《白马啸西风》	57,399	38,850	《苍穹神剑》	182,340	125,444
《射雕英雄传》	783,734	534,807	《飘香剑雨》	451,102	306,530
《倚天屠龙记》	818,875	569,529	《绝代双骄》	646,795	449,923
《鹿鼎记》	1,031,870	687,166	《陆小凤传奇》	773,958	524,308
《越女剑》	13,642	13,642	《风铃中的刀声》	131,825	91,778
共计	2,705,520	1,843,904	共计	2,186,020	1,497,983

经过初步统计,我们不难发现发现金庸的武侠小说字数更多,篇幅更长。通过对比,我们可以发现金庸的武侠小说篇幅不一,除常见的长篇小说之外,《白马啸西风》与《越女剑》分别属于中篇、短篇小说。相反,古龙更偏向于创作长篇、中篇小说,所创作的短篇武侠小说较少。

2.2. 研究工具与研究方法

本次统计主要结合使用 AntConc 与 freqci2009 两种工具。其中, AntConc 主要用来进行语境共现、搭配分析、提取关键词表等操作。在进行文本分词、统计词频时候则需要运用分词软件 freqci2009。AntConc 作为日本早稻田大学 Laurence Anthony 博士实验室研发的、基于 Microsoft Windows 操作系统的多平台的语料库分析工具包,日益受到语言学领域相关研究者的重视与推崇。作为自动分词软件, freqci2009 综合采用正向最大匹配法与逆向最大匹配法,且具备统计词频的辅助功能。综上,此两种统计工具基本能够满足本次研究的统计要求。

基于统计目的与统计工具的功能,本次统计将从常用句式、高频词、色彩词的使用情况及相关搭配特点等角度进行统计分析。我们本次是通过定量统计与定性分析相结合的方法,具体为观察统计结果,列表分析,对比研究等方式,得到关于金庸、古龙不同语言风格的定量与定性信息,进而开展系统性的语言风格分析工作。

3. 常用句式特点与比较

为了进一步研究两位作者的常用句式特征,我们通过添加通配符,设计句型分类标准,统计相同句式等操作,建立了金庸、古龙小说的句式语料库。新建的句式语料库将用于后续的统计分析工作。

关于划分句式的标准,本次主要参考了北京大学中文系现代汉语教研室版《现代汉语》关于现代汉语句式类型的相关论述[6]。另外,由于从句子所表达的内容来看,用语气分类所得到的句子更能集中反映作者意图达到的表意目的。因此,我们将从语气角度进行句类划分。最终,我们选择陈述句、疑问句、祈使句、感叹句、呼应句共计五类常用句式,并将统计结果列入下表 2:

Table 2. Statistics of sentence patterns commonly used by Jin Yong and Gu Long (in the perspective of sentence category)
表 2. 金庸、古龙常用句式使用情况(句类角度)

指标 \ 句类	金庸(频次)	频率	句类	古龙(频次)	频率
陈述句	76,122	0.72335	陈述句	54,801	0.66877
疑问句	17,147	0.16294	疑问句	17,660	0.21552
感叹句	6471	0.06149	感叹句	7346	0.08965
祈使句	3810	0.03620	祈使句	1699	0.02073
呼应句	1685	0.01601	呼应句	437	0.00533
共计	105,235	1.00000	共计	81,943	1.00000

根据统计结果，我们可以发现除了陈述句外，金庸、古龙在行文时都有偏好使用疑问句的习惯。一般而言，疑问语气的特点在于不确定性强，常用作提出问题或询问情况。在金庸、古龙的武侠小说中，其频繁出现在人物对话与叙事语言中，用以推动故事情节的发展。由此可见，正是武侠小说的行文特点造成了疑问句使用频率较高的现象。同时，疑问语气的广泛使用也营造了武侠小说矛盾激烈，故事节奏快的氛围。

在此基础上，我们将目光聚焦于金庸、古龙作品的句长及各句式的使用频率差异，尝试探究二者句式的深层差别。相关统计结果如表 3 所示：

Table 3. Statistical analysis of Jin Yong and Gu Long sentences (in the perspective of sentence category)
表 3. 金庸、古龙句长统计分析(句类角度)

统计指标 \ 作品来源	平均句长(单位：词)	频率方差(S ²)	频率标准差(s)
金庸	17.52177	0.071027	0.26651
古龙	18.28079	0.060439	0.24584

通过分析表 3 信息，我们利用绘图工具绘制了如下统计组合图(图 1)：

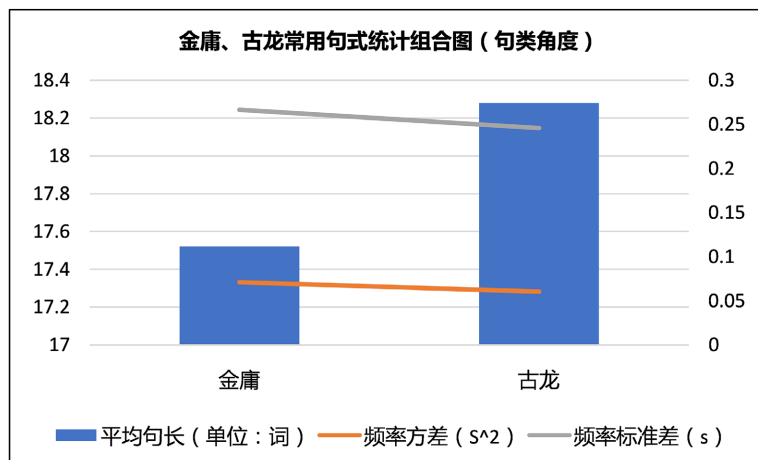


Figure 1. Statistical combination chart of common sentences of Jin Yong and Gu Long (in the perspective of sentence category)
图 1. 金庸、古龙常用句式统计组合图(句类角度)

从句长统计结果来看,二者的平均句长较为相近,而古龙平均句长略长。在武侠小说中,古龙多用形容词、副词来堆砌长句,造成句子普遍较长的现象。这是由于古龙在小说中常作分析与评论之语,而长句的形式特点更便于文学性表述的展开。另外,长句的使用更利于作者表情达意,而古龙的小说更多哲理性与文学性名句的原因也出自于此。从频率方差与标准差的统计结果来看,金庸武侠小说的句式离散程度大于古龙。这说明金庸句式更加灵活多变,不拘一格。古龙小说句式的稳定性更强,语言风格更不易改变。

其中,句长及频率方差的不同反映了语体选择,即书面语与口语使用策略的差异性。相对而言,金庸的语言体现出书面语与口语相结合的特点。相比古龙,金庸武侠小说的口语化倾向更强,例如《倚天屠龙记》中丐帮弟子以及《鹿鼎记》的主人公韦小宝多用市井俚语与方言词汇,甚至经常使用粗鄙之语。相较于古龙多用文学性、书面化语言的特点,金庸更懂得结合书面语与口语的特点,达到雅俗相融的境界。由此,金庸的著作也更易为大众所接受与喜爱。

4. 高频词分析与对比研究

高频词是反映作者用语习惯的重要指标。根据之前统计结果,我们发现金庸作品的总词数共计1,843,904个,古龙作品的总词数共计1,497,983个。在语料总体规模较大的基础上,统计高频词的使用情况能够更客观地反映作者词汇风格以及用词偏向性,进而为对比分析二者用词习惯的相似处与差异性提供可能。

在开启了识别专有名词的基础上,我们统计了金庸与古龙作品高频词汇的使用情况。由于汉字中存在大量虚词的特点,预计两位作家最高频率的词汇均属于虚词,不能如实地反映作家用语风格。因此,我们人工剔除了前20位的高频词,将第21~50位词语作为本次统计的高频词,统计情况见表4:

Table 4. Comparison statistics of Jin Yong and Gu Long's high frequency words (Top 21~50)

表 4. 金庸、古龙高频词对比统计(第 21~50 位)

排序	金庸作品高频词	出现次数	频率	排序	古龙作品高频词	出现次数	频率
21	说	7601	0.0041222	21	却	7450	0.0049734
22	去	7339	0.0039801	22	陆小凤	7141	0.0047671
23	无	7242	0.0039275	23	来	7137	0.0047644
24	两	7219	0.0039151	24	要	6841	0.0045668
25	韦小宝	6774	0.0036737	25	小	6775	0.0045227
26	要	6472	0.0035099	26	得	6711	0.0044800
27	中	6447	0.0034964	27	没有	6649	0.0044386
28	大	6372	0.0034557	28	那	6547	0.0043705
29	着	6201	0.0033630	29	个	6207	0.0041436
30	张	6072	0.0032930	30	说	5696	0.0038024
31	便	5948	0.0032258	31	知道	5105	0.0034079
32	都	5924	0.0032127	32	儿	4979	0.0033238
33	已	5840	0.0031672	33	两	4938	0.0032964
34	将	5813	0.0031526	34	还	4882	0.0032590

Continued

35	个	5721	0.0031027	35	去	4822	0.0032190
36	只	5470	0.0029665	36	但	4769	0.0031836
37	见	5392	0.0029242	37	鱼	4565	0.0030474
38	上	5382	0.0029188	38	什么	4560	0.0030441
39	却	5230	0.0028364	39	看	4204	0.0028064
40	说道	5209	0.0028250	40	很	4142	0.0027651
41	忌	5122	0.0027778	41	想	4087	0.0027283
42	和	5090	0.0027604	42	自己	3925	0.0026202
43	三	4678	0.0025370	43	笑	3879	0.0025895
44	但	4617	0.0025039	44	他们	3755	0.0025067
45	小	4414	0.0023938	45	中	3699	0.0024693
46	郭靖	4111	0.0022295	46	和	3620	0.0024166
47	好	4041	0.0021915	47	只	3616	0.0024139
48	就	3807	0.0020646	48	上	3594	0.0023992
49	给	3781	0.0020505	49	会	3541	0.0023638
50	自己	3762	0.0020402	50	种	3513	0.0023452

除大量使用的人物姓名外，我们能从对比情况中发现金庸、古龙善于描绘江湖意气的场面。具体而言，“笑”、“来”、“去”等高频动词的大量使用充分体现了快意恩仇的江湖气魄。两位作家的文本中，“笑”常搭配出现“大笑”、“冷笑”、“含笑”等表述。“来”“去”的频繁出现反映出侠客们身形自如，潇洒不羁的形象特征。

通过删去重合的高频词，我们发现金庸小说中特有高频词为 17 个，古龙小说的特有高频词也为 17 个，独特词高频数量完全相等。具体而言，金庸的特有高频词有“大”(频次 6372，频率 0.0034557)、“三”(频次 4678，频率 0.002537)、“好”(频次 4041，频率 0.0021915)等。结合文本中词汇的搭配与使用情况，我们发现金庸笔下的武林好手与武功招式多以“三”命名，例如《白马啸西风》中的“吕梁三杰”、“白马李三”，《倚天屠龙记》中的“张三丰”、“昆仑三圣”与“三才剑阵”等，均体现出作者对数字“三”的偏爱。

另外，“中”(频次 6447，频率 0.0034964)成为高频词汇之一与“中国”(频次 445)、“中原”(频次 230)、“中土”(频次 135)等词汇占比较高的情况密不可分。根据上述常用搭配，我们可以发现金庸武侠小说多有具体的叙事空间，侧重以古代中国地理与历史文化作为背景。这反映出金庸具有自觉运用中国传统文化的强烈意识，笔下的文字散发着中国传统文化的浓郁气息。

相比而言，古龙的独特词为“自己”(频次 3925，频率 0.0026202)、“没有”(频次 6649，频率 0.004438635)、“很”(频次 4142，频率 0.002765051)等。通过搜索相应词汇的搭配情况，我们发现在古龙小说中，“自己”常出现于侠客心理描写与作者的评论文字当中，反映了侠客微妙的心理活动以及孤身孑然的特点。类似地，“没有”一词也多由古龙用来营造寂寞空旷的叙事环境。其中，“没有”常搭配用作“没有人”(频次 494)、“没有说”(频次 262)等表述，表现侠士无人匹敌或无人知心的孤独处境。正合张志岩(2019) [7]所描述“侠客们身上的虚无色彩与孤独内蕴”的气质特点。相比于金庸恢弘的叙事空间与深厚的文化

底蕴，古龙倾向于临摹人的情感世界，以及描绘个体的生活体验。

值得注意的是，古龙小说中“他们”(频次 3755, 频率 0.0025067)也是一个高频词。然而，“她们”(频次 479, 频率 0.0003197)的频率明显偏低，不能排入前 50 位高频词的序列。“他们”的广泛使用充分体现古龙的武侠世界中侧重描绘男性群体，重要人物以男性为主，男性气质浓厚的独特之处。

5. 色彩词使用情况对比与分析

5.1. 理论基础

颜色偏好是人在心理上对某些颜色的偏爱性。心理学认为，颜色的偏好影响了选取也在幽微之处显示出人与人之间的个体差异。近年来，随着色彩心理学的发展，学界对于颜色与人物性格及心理的研究更加深入。Palmer 和 Scholss (2010)对颜色偏好提出了一种生态价效理论(The Ecological Valence Theory, EVT)。他们认为，人们之所以会对某种颜色产生偏好，是因为其对与该颜色相关联的物体或实体存在好感，并系统佐证了颜色与人类知觉感情，思想意志的紧密联系[8]。

因此，我们可以将颜色词频率的高低作为判断不同文本差异度大小的依据，进而判断作者是否具有相似的语言风格。

5.2. 统计过程及结果分析

为了增强统计结果的精确性与可信度，我们在统计之初便考虑到汉语中专有名称或惯用词组中可能带有颜色字词的情况。因此，我们在统计之前便删去了的“明白”、“青春”、“青州”、“黄河”、“黄泉”等带色彩的常用字样，尽力使得统计结果更能体现作者本身的颜色偏好与使用情况。

最终，我们将金庸、古龙作品中色彩词的使用情况列入表 5:

Table 5. Statistics on the use of color words in Jin Yong's and Gu Long's works

表 5. 金庸、古龙作品色彩词使用情况统计

金庸作品色彩词	频次	频率	古龙作品色彩词	频次	频率
黄	5847	0.0031710	白	3223	0.0021516
白	2871	0.0015570	黑	1872	0.0012497
青	1893	0.0010266	青	1395	0.0009313
红	1521	0.0008249	红	1344	0.0008972
黑	1319	0.0007153	黄	761	0.0005080
绿	238	0.0001291	绿	366	0.0002443
蓝	137	0.0000743	紫	361	0.0002410
紫	147	0.0000797	蓝	357	0.0002383
橙	2	0.0000011	橙	0	0.0000000
共计	13,975	0.0075790	共计	9679	0.0064614

根据统计结果，从高频颜色词来看，金庸最常用的前三色彩词为“黄”(5896 次，频率 0.00317)、 “白”(2871 次，频率 0.00156)、 “青”(1893 次，频率 0.00103)。我们可以看出，金庸酷爱“黄”一词。作为金庸最喜爱的色彩词，金庸武侠小说中大量主要角色的名字中都带有“黄”。例如《射雕英雄传》中的黄蓉、黄药师、黄裳，《倚天屠龙记》中的黄衫女等。在金庸的文本中，“黄”、“红”等表示明艳、

亮丽等暖色调的色彩词出现频率较高，而“黑”、“绿”、“蓝”等表示黯淡的冷色调色彩词出现较少。

古龙最常用的前三色彩词为“白”(3223次，频率0.00215)、“黑”(1872次，频率0.00125)、“青”(1395次，频率0.00093)。根据文本中体现的常用搭配，古龙偏爱“白衣”(229次)、“苍白”(229次)、“雪白”(59次)等表述。除此之外，古龙还冠以许多人物“白”之姓名，例如“许白”(249次)、“白凌霄”(21次)。古龙对“白”情有独钟的独特现象，充分体现其对白衣翩翩公子形象的迷恋。同样地，古龙也偏好“黑”色。在古龙的武侠小说中，“黑”常搭配为“黑衣”(频次411)、“黑暗”(364次)等，以表示人物活动、事件进展的背景信息及叙事元素。

通过对比，我们不难发现金庸武侠小说整体上属于偏好鲜艳明朗、光芒四射的色彩基调，带给读者温暖亮丽的氛围体验。古龙的武侠小说“绿”、“紫”、“蓝”等冷色调颜色使用频率更高，同时更注重“黑”、“白”反义颜色词的使用也有助于烘托强烈对比之下的反差感。总体而言，金庸偏爱使用明媚绚烂的色彩来烘托温暖和煦的叙事氛围，古龙则偏向于营造一种冷峻严肃的压抑之感。

6. 结论与展望

6.1. 结语

经过上述统计分析，我们可以发现金庸与古龙的语言风格既有相似之处，也存在着明显的差异。我们现从相似性与差异性两个方面来总结本次统计分析发现的结果。

6.1.1. 语言风格的相似性对比

综合来看，二者的相似性主要从分析高频词的角度得出。通过统计，我们发现两位作家都大量利用“笑”、“来”、“去”等词汇。两位作家在书写武侠小说时，都侧重于描绘武林世界的自由感与洒脱感，创作出多个潇洒不羁、快意恩仇的侠客形象。

6.1.2. 语言风格的差异性对比

相较而言，两位作家的不同之处更为突出，主要表现为常用句式、高频词汇以及色彩词各方面的差异。

句式层面，古龙习惯使用长句进行表述，其语言更富有书面化色彩。金庸则将汉语口语与书面语进行融合，既发挥了书面语长于议论与白描的优势，也利用了口语灵活性强、变化性丰富的特点，造成句式风格更加灵活多样的特色。

高频词方面，金庸习惯使用“三”、“中”等词，其高频词的搭配情况充分体现了作者受中国传统历史与文化影响之深。古龙则偏爱“没有”、“自己”、“他们”等词，着重表现侠客睥睨天下、寂寞孤傲的高冷气场以及男性气质浓厚的特征。

色彩词方面，金庸明显偏爱“黄”、“红”等明艳的暖色调，多借此营造温暖宜人、哀而不伤的语言环境。古龙则明显痴迷于“白”的使用，且多辅以“黑”之色彩，造成强烈的色彩反差效果。

总体而言，金庸的语言很大程度上受传统文化的浸润，更加体现了家国一体，天下为家的伟大理想。与之相对，古龙的语言受个人的雕琢感更强，着重体现人本身的精神风貌与形象魅力。正是二者语言风格的出入，造成了金庸善写大侠，古龙善写公子的不同特点。

6.2. 未来研究方向

本次研究更多基于常用句式、高频词、色彩词的统计分析。受限于统计工具与统计方法的不足，本文较少涉及语义及篇章层次的计量分析。然而，这一领域的相对空白的现状也说明此类研究具有广阔的发展空间与探究价值。我们相信，随着统计技术的发展与不断完善，未来的相关学者将会立足于语义、

段落、语篇等领域，利用计量方法深入研究作家语言风格与文本特点，最终得出关于作家语篇特点的整体性结论。

类似地，分词工具同样有较大的发展空间。许多中国作家对于模仿古人用语习惯以遣词造句的兴趣颇丰。金庸、古龙的小说文本中也存在多处使用古诗词与古代文言文的现象。然而，本次研究使用的分词软件 `freqci2009` 以及至今所开发的大多数统计软件更多是基于现代汉语信息处理的特点所研发的，在处理古文、诗句方面的效果明显不尽人意。未来，倘若能研制出配套处理古代汉语的统计工具，将会对分析此类带有古汉语字句的作品提供更全面的帮助，提高分词的正确率与分析的准确性。

参考文献

- [1] 徐岱. 论金庸小说的艺术价值[J]. 文艺理论研究, 1998(4): 15-28.
- [2] 芦海英. 论古龙武侠小说的现代性特征[J]. 淮海工学院学报(社会科学版), 2006, 4(1): 43-46.
- [3] 王开银. 金庸、古龙武侠小说语言风格比较研究[D]: [硕士学位论文]. 乌鲁木齐: 新疆大学, 2008.
- [4] 胡菁娜. 金庸与古龙武侠小说创作风格的比较[J]. 龙岩学院学报, 2013, 31(6): 98-101.
- [5] 赵铭. 金庸古龙小说文化异同的语境原因分析[J]. 明日风尚, 2016(17): 389.
- [6] 北大中文系现代汉语教研室. 现代汉语[M]. 北京: 商务印书馆, 2018.
- [7] 张志岩. 论古龙武侠小说的现代理念与创作实践[D]: [硕士学位论文]. 沈阳: 辽宁大学, 2019.
- [8] Palmer, S.E. and Schloss, K.B. (2010) An Ecological Valence Theory of Human Color Preference. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 8877-8882.
<https://doi.org/10.1073/pnas.0906172107>