余华与莫言长篇小说的计量统计和分析*

涂梦纯[1]，刘颖[1]

（1.清华大学人文学院中国语言文学系，北京 100084）

摘要：该文以余华和莫言的各 5 部长篇小说作为语料，选用段落、句子和分句的长度，标点及其 n 元，词类及其 n 元，词汇及其 n 元和颜色词作为特征，使用层次聚类、k-s 检验从整体上判断两位作者的作品是否有差异，再使用秩和检验判断两位作者的某一具体特征是否有差异。该文对余华和莫言用词上的特征进行了详尽的分析，发现了莫言用词丰富、情感充沛及文言化、乡土化的特征，而余华与之相比白话、冷静、讽刺的风格。

关键词：计量风格；余华；莫言；假设检验；层次聚类

中图分类号：TP391　　　　**文献标识码：**A

Statistics and Analyses of Yu Hua and Mo Yan's Long Novels

Mengchun Tu[1], Ying Liu[1]

(1. Department of Chinese Language and Literature, Tsinghua University, Beijing, China, 100084)

Abstract: This article uses the five novels of Yu Hua and Mo Yan as the corpus. The lengths of the paragraphs, sentences and clauses, punctuation, part of speech, words and their n-gram and color words were selected as the features. As for Statistical method, Clustering and KS test were used to overall judge whether there were differences between the two authors, and then Wilcoxon test were used to jugde whether there is a difference between a specific characteristic of the two authors. This article makes a detailed analysis of the characteristics of Yu Hua and Mo Yan's words, and finds that Mo Yan has abundant words, strong emotions, classical and regionalisms. Relatively, Yu Hua shows a calm and satirical style.

Key words: quantitative stylistics; Yu Hua; Mo Yan; hypothesis test; hierarchical clustering

1 引言

余华，1960 年生于浙江杭州，先后创作了《在细雨中呼喊》《活着》《许三观卖血记》《兄弟》《第七天》等 5 部长篇小说和大量短篇、随笔、杂文；莫言，1955 年出生于山东高密，先后发表了《红高粱家族》《酒国》《丰乳肥臀》《生死疲劳》《蛙》等 11 部长篇小说、27 部中篇小说和若干短篇、散文和电影剧本。两人出生年代相近、创作主题相似，且都是当代风格鲜明、享誉中外的作家，莫言更于 2012 年荣获诺贝尔文学奖。

已经有很多学者对比过他们在创作上的同异，徐文明[1]指出余华的暴力叙述冷静而职业化，准确而尖锐，而莫言的语言则富有感性和想象力，使文章整体呈现出一种宏大叙事和慷慨赴义的英雄主义情结；李艳军[2]对比余华的《活着》和莫言的《丰臀肥乳》，说前者饱含悲凉而痛苦，叙事方式又显得温和而淡定，而后者气势宏大、洋洋洒洒、浓墨重彩；毛冰[3]则说，余华内敛克制、朴素中饱含张力，莫言想象大胆乃至恣意妄为、爆发力强。除了主题思想以外，还有一些语言学方面的研究，如王清溪[4]、王委艳[5]都指出了余华在小说语言上的陌生化实验，季红珍[6]和杨连芬[7]也说明了莫言写作时在词汇、修辞等方面的陌生化特征。

检视前人的研究我们发现，对两人的对比研究主要集中在作品内容方面，对语言风格的描述或者是主观的、或者是独立的，对二者语言风格的客观对比目前还是一片空白。因此，本文希望从计量风格学的角度对两人的文学语言进行研究。

计量风格学区别于传统风格学，不以个人阅读体验为依据，而转向客观方法来分析写作风格。在中文文本的领域，通常会提取长度层面的词长、分句长、句长、段长[8]，词汇层面的标点、词类、词汇丰富度、虚词、高频词[9]等，语音层面的元音和辅音、 节奏等[10]，以及 n 元文法等方面的特征。其中 n 元文法又可以选择标点、词汇、词类、语音[11]等多个方面的特征进行提取。再通过统计学方法来判断不同作者之间是否存在差异，最后分析这些数据体现了文本怎样的风格差异。

---

* **收稿日期：**　　　　　**定稿日期：**

基金项目：北京市社科基金项目“莫言与余华小说的计量风格分析与比较”（16YYBD021）；教育部人文与社科项目“语体特征的自动提取和研究”（17YJAZH056）

作者简介：涂梦纯（1992——），女，硕士，主要研究方向为语料库语言学；刘颖（1969——），女，博士，教授，主要研究方向为语料库语言学和计算语言学。

2 语料库和预处理

本文选择余华全部的5部长篇小说及莫言与余华作品发表年份相近的5部长篇小说为研究语料，并为了在图表中年方便表示重新编号，具体文本及对应编号见表1。首先对文本进行预处理：删去章节号、标题、前言与后记及《蛙》后的同名剧本。再使用NLPIR汉语分词系统[1]对文本进行分词，并据此统计语料库中字、词、标点的数量和分句、句子和段落的长度，前者见表1，后者情况见表2。

表1 语料库字、词、标点情况

编号	书名	余华 字数	词数	标点	编号	书名	莫言 字数	词数	标点
Y1	呼喊与细雨	130051	100735	13148	M1	红高粱家族	235153	200242	32817
Y2	活着	74910	68176	11313	M2	酒国	172685	143992	23797
Y3	许三观卖血记	104957	94101	17266	M3	丰乳肥臀	404372	344302	57626
Y4	兄弟	350673	279328	42619	M4	生死疲劳	335963	283207	46822
Y5	第七天	84689	69938	11502	M5	蛙	139594	116839	19265
合		745280	612278	95848	合		1287767	1088582	180327

将选择10部小说按作者和时间编号为Y1-Y5和M1-M5,在之后的图表中也按此标注，不再重复说明。

3 段落、句子和分句

在预处理时已经统计所有10部小说的段落、句子和分句的数量及长度，现将其数量、均值、最大值和离散度（即标准差）列入表2。

表2 语料库段落、句子及分句情况

	编号	Y1	Y2	Y3	Y4	Y5	M1	M2	M3	M4	M5
段落	数量	2212	2046	2172	5038	1692	3173	2361	3579	3091	1836
	均值	58.79	36.61	48.32	69.61	50.05	74.11	73.14	112.98	108.69	76.03
	最大值	553	402	623	705	487	1557	3360	1464	2259	2251
	离散度	65.48	45.22	52.93	80.39	66.02	111.38	153.71	138.22	167.02	128.08
句子	数量	4909	3263	3577	10089	3097	9142	7286	16586	12467	6003
	均值	26.49	22.96	29.34	34.76	27.35	25.72	23.7	24.38	26.95	23.25
	最大值	105	149	230	307	168	303	258	478	309	230
	离散度	15.48	16.94	24.86	25.15	21.28	19.15	20.06	17.29	20.37	17.38
分句	数量	11504	9324	13073	34112	8898	27440	20277	49456	40540	18160
	均值	11.3	8.03	8.03	10.28	9.52	8.57	8.52	8.18	8.29	7.69
	最大值	28	33	30	36	41	76	96	53	61	41
	离散度	4.86	3.86	3.91	4.8	5.02	5.22	5.34	4.79	4.91	4.41

通过数据能够发现，在最大值上，莫言不论是段落、句子还是分句都远超余华；在均值上，莫言的平均段长大于余华，而平均句长、分句长则小于余华。莫言平均段长大于余华的原因是莫言段落中更多的句子。莫言这种更长的段落降低了文本的易读性，需要读者花费更多的精力在文本阅读上，也给文本增加了更多可能性；但莫言文本的段落参差不齐，读起来长短不一、错落有致，能够有效的缓解读者长时间阅读长段落的疲惫感。

另一方面，将同一作者5部作品的长度分布列整合后，对余华和莫言的段长、句长和分句长进行K-S检验（Kolmogorov–Smirnov test）[12]159-167。该检验的核心思路是将原分布列

1 又名ICTCLAS2016，下载网址 http://ictclas.nlpir.org/

转换为经验分布函数，再通过两条分布列差的上确界与参考值进行对比判断原分布是否相同。输出结果中段长和句长的 p 值小于 0.05 而分句长的 p 值大于 0.05，即余华和莫言在段长和句长的分布列上存在显著差异。

4 标点

4.1 层次聚类

层次聚类可以对选定特征的整体进行比较，以 n 元文法对语料库中的标点信息进行聚类，结果如图 1。

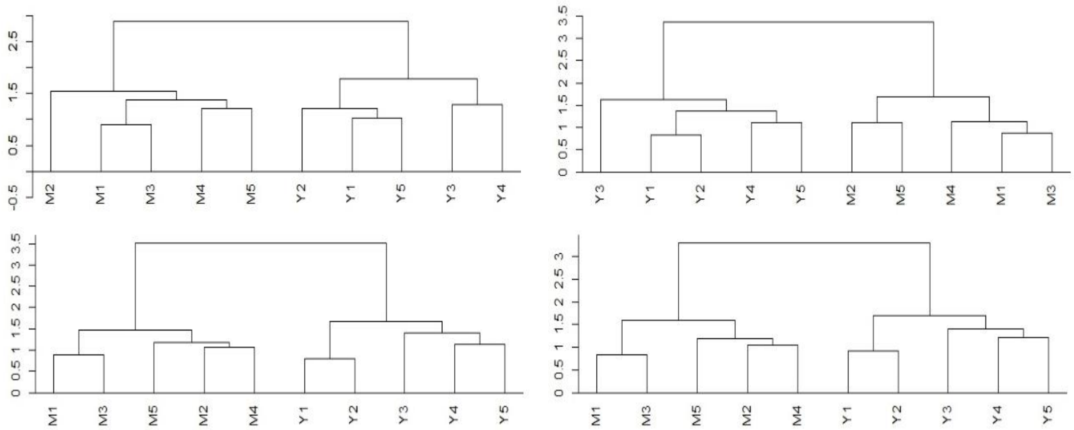


图 1 语料库基于标点的 1-4 元聚类

本文在进行层次聚类时，文本间的距离使用欧氏距离计算，类与类之间的距离使用离差平方和（ward）计算，以下不再特意说明。可以看到在 n=1-4 时，10 部小说按作者聚成两类，表示同一作者文本之间的距离均小于不同作者文本之间的距离。以最上方 n=1 时为例，莫言的文本在纵坐标 1.5 处聚成一类，余华的文本则在纵坐标 2.0 左右聚成一类，但余华和莫言的文本则在纵坐标 3 左右才相聚。以上结果说明，标点频率本身及标点的多元组合是能够区分余华和莫言写作风格的特征量。而具体哪些标点的使用令余华和莫言的文本区分开来，则需要进行进一步的研究。

4.2 秩和检验

使用秩和检验来判断两位作者在哪些标点的使用上存在显著差异，且哪方在该标点上的频率占显著优势：检验结果小于等于 17 时，该标点莫言的使用频率显著大于余华，大于等于 38 时则相反。秩和检验的结果如图 2，可见莫言在叹号、顿号、括号、破折号上的使用频率显著大于余华，不存在余华使用频率显著大于莫言的标点。

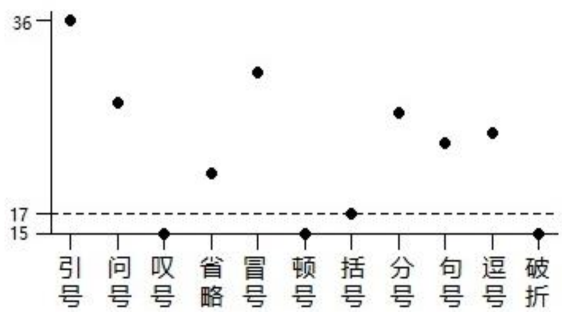


图 2 余华、莫言 10 部小说标点的秩和检验

4.3 叹号、顿号、括号和破折号

感叹号用以表示强烈的感情。

例 1 余占鳌从劈柴堆里跳起来，手舞足蹈地大喊：“是老子的！哈哈！是老子的！”（摘自《红高粱家族》）

莫言多于余华的感叹号用量代表着前者文本中更为强烈而充沛的情绪。

顿号表示并列的词或词组之间的停顿。

例 2 桥洞、桥墩、钢梁、铁架、狗皮大衣狐皮帽子，杏黄爬犁蒙古马，铁路桥周围的一切都纤毫毕现，连一根毛掉在冰上都能看得清清楚楚。(摘自《丰乳肥臀》)

顿号不论是用在名词、形容词、动词还是其他词类之间，都代表了莫言对某一场景或事件更丰富的描述。

括号是文中注释部分所用的符号。

例 3 粮库保管员肖上唇的儿子肖下唇（后来改名为肖夏春）则说，中国的飞行员吃得比苏联飞行员还要好。(摘自《蛙》)

括号的整体用例比感叹号和顿号更少，在余华的文本中仅出现 10 例，而莫言的文本中共 63 例，这种用于解释和补充说明的符号的显著优势，体现了莫言小说中对于人物更加具体的刻画。

破折号的用法相对较多。首先是拉长语音的功能。

例 4 高粱红了——日本来了——同胞们准备好——开始开炮——

与同样表达强烈语气的感叹号相比，给人以更加嘹亮、开阔的感觉。破折号所表达的更具画面感，好像能听到长长的口号在广袤的红色高粱地里回荡。意象与语音一起延长，更能调动读者情绪和想象。

其次是与括号类似的解释说明的功能。

例 5 老县委书记金边——就是那位为我爹的黑驴挂过铁掌的小铁匠——升任市人大副主任，庞抗美接班已成定局。(摘自《生死疲劳》)

虽然功能类似，但破折号与括号仍有不同。被破折号夹在中间的文字属于正文，而括号中间的文字只是注释，因此从客观数据的角度来说，破折号造成了语气的停顿，将句子打断成几个分句，这一现象映射到主观的阅读体验上，就会因句子没有结束而让读者一直提着一口气，将目光继续移动下去，在另一种层面上达到了延展声音的效果。

最后是话题或对象的转化的功能。

例 6：嫂子，要想跟你亲个嘴，必须踏着梯子！——五官，你给我闭嘴！ (摘自《蛙》)

在《蛙》中出现了大量不使用引号而用破折号转换说话人的方式，不使用正反引号来打断文字，而是用破折号来进行对话者的迁移，仍然隐约体现了其拉长语音的功能：两个人的对话紧紧连成一片，让“我母亲”的话音更大，去打断五官的调戏。这种丰富的修辞作用是简单的引号不能代替的——而破折号的这种使用方法也是仅存在于莫言的文本中的。

5 词类和词汇的定量分析

5.1 层次聚类

同样首先使用层次聚类的方式对余华和莫言小说中的词类和词汇进行整体判断，结果见

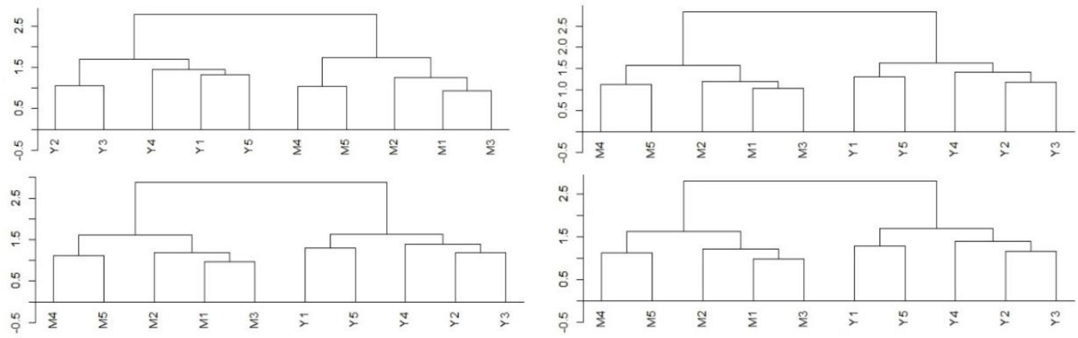
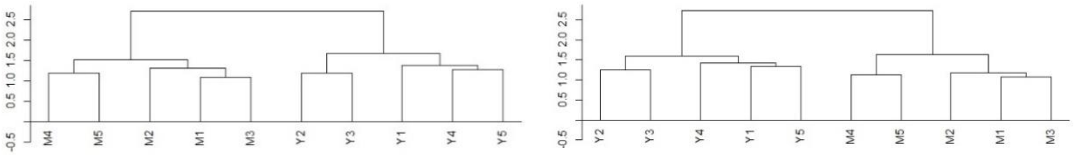


图 3、图 4。

图 3 语料库基于词类的 1-4 元聚类

图 4 语料库基于词汇的 1-2 元聚类

n=1-4 时，10 部小说按作者聚成两类，说明词类频率（n=1）及词类的 2-4 元组合是能够区分余华和莫言写作风格的特征量。而以词汇作为特征、在 n=1 和 2 时，10 部小说按作



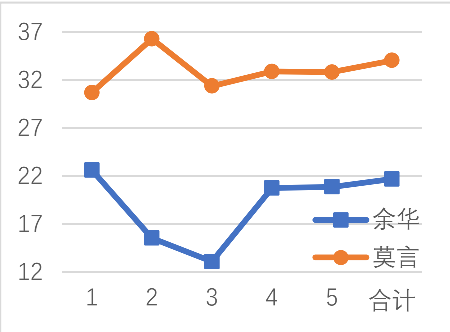
者分成两类。说明两位作者在词汇和词汇的二元组合（可以认为是词组）上的确存在显著差异。

5.2 词汇丰富度

词汇丰富度是描述文本中作者用词的重要维度，主要类符形符比（Type-Token Ratio，TTR）来表示。所谓类符（type）是文本中所有出现的词汇的类数，相同的词记为 1 个；形符（token）是文本总词数。但是，通常一个作者在写作时使用的不同词汇数量是一定的，而随着文本的增长，总词数是一定在增加的，所以文本越长、对 TTR 的结果影响越大。因此，本文采用 Guitaud[13]提出的修正公式，取形符的平方根作为分布计算词汇丰富度：

$$\sqrt{TTR} = Types/\sqrt{Tokens} \tag{1}$$

为了便于观察，将由此得到的余华、莫言修正类符形符比绘入图 5。可以看到两条折线完全分开，且莫言的词汇丰富度明显大于余华。因此，需要寻找余华和莫言具体的词汇差异，使用秩和检验进行。



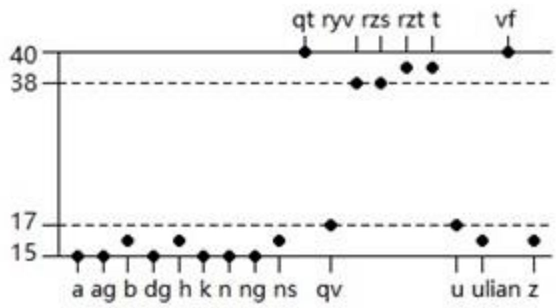


图 5 语料库的根号词汇丰富度　　　　图 6 语料库词类的秩和检验

词类的使用频率差异能够反映两位作者写作风格的一些差异。例如，名词代表的是小说世界中出现的人或物，是被描写和记叙的对象，莫言明显大于余华的使用频率代表着前者小说世界中更加丰富的物象；而形容词、区别词和状态词都是修饰成分，用来描述小说中所提及的对象，两位作者之间显著差异的使用频率其实是与名词相辅相成的，当然也意味着莫言文本空间中更加丰富多彩的形象。

再对词汇进行秩和检验。首先，莫言使用而余华未使用的词汇，称为莫言的“独现词”；余华使用而莫言未使用的词汇，称为余华的“独现词。”其次，对于莫言和余华均有使用的词汇，秩和检验又将之分为 3 类：

第 1 类，频率无显著差异的词汇，共有 12887 个；

第 2 类，频率存在显著差异且余华的使用频率高于莫言的词汇，共 257 个；

第 3 类，频率存在显著差异且莫言的使用频率高于余华的词汇，共 1263 个。

将第 2 类词与余华的独现词合并，可以认为这 4069 个词是能够反映余华文本特征的特色词；同理，将第 3 类词语莫言的独现词合并，这 22610 个词就是莫言的特色词。分类情况如图 7。

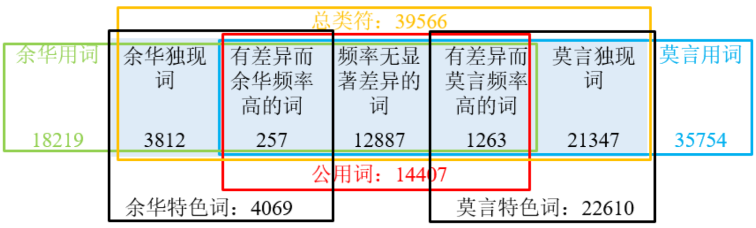


图 7　语料库的词表分类情况

第六章的具体风格分析，就围绕着以上词类和词性的秩和检验结果展开。

6 用词的风格分析

6.1 莫言丰富的类符和余华单调的类符

在秩和检验中表现出频率显著性的词类如名词（n）自不必说，即使在检验中两位作者频率无显著差异的词类上，莫言也表现出了丰富的词汇选择空间，如拟声词和量词，两者对

比类符对比详见表 3。还需说明表 3 种的几个表达方式：用下划线标注的词汇有重叠式，如“嗖”有“嗖”和“嗖嗖”两种形式；用斜杠标明的如“吱吱扭扭\吜吜”则代表有“吱吱扭扭”和“吱吱吜吜”两种形式。

表 3 余华、莫言特色词中量词和拟声词的对比

词类	余华	莫言
量词	付、篮、滩、公分、平米、回合、毫米、句、些、幢	柄、杆、缕、瓢、响、厘米、挺、簇、栋、篓、发、绺、勺、汪、摞、股子、尾、抹、瓣、坨、筐、针、公斤、捧、梭、手、椇、侧、穗、马力、顷、瓦、袋子、味、伏、筒、成、千斤、罐、丘、克、磅、坛、环、碟、码、挂、帖、炷、柱、石、重、则、海里、周年、升、棚、爿、周岁、票、华里、剂、锹、℃、池、帘、档、厘、伏特、房、平方米、峰、平方公里、公尺、美金、口袋、畦、出、穴、兜、炉、根、匹、支、尺、批、株、枝、枚、窝、腔、线、帮、股、顶、部、粒、项、条、只、种、面、盏、滴、行、盆、寸、幅、丈、堵、桩、所
拟声词	哇哇、嘿嘿、吧哒吧哒、怦怦、乒乒、突突、叽叽喳喳、哇噢哇噢、哼哼哈哈	乒、乓、春、嘭、嗖、噌、梆、叭叭、哞哞、哞哞哈哈、咕噜噜、嚓、嚓嚓、哼哼唧唧、刮刮唧唧、叽叽咕咕、噗噗、噗噗哧哧、噗通、噗哧噗哧、哔哔叭叭、哔哔剥剥、哗啦、哗啦啦、哗哗啦啦、啪嗒、呱嗒、吧嗒、嗒嗒、嘀嘀\滴滴嗒嗒、滴滴答答、滴溜溜、哒哒、啪哒啪哒、咔哒、咔哒咔哒、咔咔嚓嚓、咕噜、咕嘟、咕咕、嘀嘀咕咕、嘎嘎咕咕、嘟嘟、咕嘟嘟、呱唧、咿咿呀呀、当当、叮叮\丁丁当当、劈啪、噼噼叭叭\啪啪、劈里\哩啪啦、唧唧喳喳、嘁嘁喳喳、嘁嘁嚓嚓、轰隆、轰隆隆、轰轰隆隆、轰隆轰隆、嘤嘤、嗡嗡嘤嘤、咯吱咯吱、咯咯吱吱、嘎嘎吱吱、吱吱扭扭\吜吜、吱吱悠悠、铛铛、哞哞哈哈、吧嗒、咣咣、咣啷、咣啷啷、咣啷咣啷、咣咣当当、咣咣采采、当啷、铿铿锵锵、啾啾、叽哩咕噜\喀喳、叽里喀喳\欻啦、哔叽、哇啦、呜哩哇啦、嘭嘭、喳、唧喳、喀嚓、哼哧、喵、喵呜、吭吭哧哧、呼啦啦、呼啦、吆儿吆儿、扑扑嗵嗵、嗵嗵嗵、呼哧、呼呼哧哧、哐啷啷啷、吱哇、嘎嘎唧唧、嘎嘎嘣嘣、呜呜噜噜、呜呜呜、嘿嘿嘿、沙沙沙、噗簌簌、簌簌、飕飕、橐橐、汩汩、淙淙、索索、得得、采采、格登、啁啾、淅沥、淅淅沥沥

5.3 中的秩和检验显示，余华和莫言拟声词和量词的整体频率没有显著差异，但两者具体使用的类符数量却如此悬殊。例如拟声词中，莫言的“丁丁当当”和“叮叮当当”、“嘀嘀嗒嗒”和“滴滴嗒嗒”和“嘀嘀答答”、“嘁嘁嚓嚓”和“嘁嘁喳喳”、“吱吱扭扭”和“吱吱吜吜”、“劈里啪啦”和“劈哩啪啦”（以及作为无差异词的“噼里啪啦”）是同音的拟声词，但他在写作时也不会拘泥于一种——或者说，莫言在拟声词的选择上本没有什么定法，而是随心所欲的选择词汇或者字眼，体现出一种非常独特的随性与灵活，比起余华的“循规蹈矩”更加自由奔放。同时，更丰富的类符选择体现了莫言对于声音的细致描摹，这让他笔下的场面更加鲜活立体，仿佛近在眼前。

此外，在表示年份的时间词上，莫言表现出他的丰富性：独属于莫言的年份——将数字、汉字和缩写合并为一个（如 1926 年、一九二六年和二六年），仅统计小说中所涉及的年份——就有 57 个之多，横跨了 1917 年至 2005 年近一个世纪的时间；而独属于余华的年份词仅有 4 个。另一方面，“早晨”“晚上”“下午”“中午”等非常普通的时间词是余华的特色词，却在莫言的文本中很少出现。据此可以认为，莫言小说的时间跨度更长——或者说莫言在写作时有意无意的强调这种跨度，而余华则与之相反，强调的是一天之内的早晚变化：这类词汇频率最高的文本并不是余华仅跨度 7 天的小说《第七天》，可以一定程度的排除文本时间本身的影响。

莫言更加丰富的词汇选择也出现在了名词上，尤其是人名和地名：

表 4 余华、莫言特色词中人名和地名的对比

名词	余华	莫言
人名	福尔摩斯、曹雪芹、川端康成、三岛由纪夫、荷马、齐白石、邓丽君、迈克尔·杰克逊、叶利钦、卡拉扬、普金	唐吉诃德、桑丘、塞万提斯、肖洛霍夫、葛利高里、婀克西妮娅、安娜·卡列尼娜、托洛茨基、亨利·基辛格、狄奥尼苏斯、巴甫洛夫、希特勒、斯大林、尼克松、高尔基、帕瓦罗蒂、白求恩、丘吉尔、拉玛西斯、奥西里斯、弗拉基米尔·伊里奇·列宁、赫鲁晓夫
地名	闽南、绍兴、江西、江苏、珠穆朗玛峰、东京、冲绳岛、大阪、南非、苏丹、开罗、海牙、塞尔特、马德里、克里姆林宫、格林威治	内蒙古、蒙古、河南、河北、沂蒙山、山东省、济南、渤海、北海、青岛、哈尔滨、烟台、威海、胶州、胶东、胶南、西安、长白山、泰山、关东、深圳、山西、黄海、湖南、东海、南海、南洋、保定、唐山、贵州、大连、香山、卢沟桥、云南、天津、长沙、莱阳、杭州、伊犁、汉口、延安、天山、大庆、莱州市、昆山、大理、四川、浙江、兰州、巫山、桂林、苏州、太行山、沈阳、塔里木、北戴河、华北、昌邑、安阳、赤水、通县、中原、山海关、豫西、上海滩、喜马拉雅、西湖、龙山、武夷山、洛阳、粤西、齐鲁、北平、金门、沙家浜、临沂、北京、龙口、西藏、徐州、秦川、阳关、曲阜、潼关、湘江、黑龙江、鸭绿江、浏阳河、雅鲁藏布江、乌江、牡丹江、乌苏里江、中华人民共和国、亚洲、南美、札幌、牛津、爪哇、波斯、英国、北极圈、满洲里、西欧、缅甸、伊豆、斯堪的纳维亚、里昂、科隆、印尼、罗马、泰国、瑞典、捷克、埃及、美索不达米亚、东南亚、菲律宾、巴西、挪威、印度、匈牙利、新西兰、大西洋、澳大利亚、越南、京都、荷兰、墨尔本、洛杉矶、柏林、美洲、英格兰、埃塞俄比亚、拉斯维加斯、伦敦、马来西亚、丹麦、日内瓦、印度洋、吕宋岛、旧金山、南美洲、苏联、西班牙、哥萨克、希伯来

余华和莫言在文中对人名和地名的引用，很多时候并不是因为要叙述这个人物、这个地方，例如：

例 7 刘作家不愧是我们刘镇的文豪，他一句话就把古今中外的作家诗人全拉过去做了自己的垫背……“连**鲁迅**先生也跟着丢脸啦，还有**李白杜甫**先生，还有**屈原**先生，屈先生爱国而投江自尽，也跟着赵诗人丢脸……还有外国的，**托尔斯泰**先生，**莎士比亚**先生，更远的**但丁**先生，**荷马**先生……多少个英名先生啊，全跟着赵诗人丢脸啦!”（摘自《兄弟》）

这些作家的名字在富有时代感和乡土气的故事里表现出突兀的文化感，在因描写底层人物而较为粗鄙的小说空间格外显眼；同时又因为略显粗鄙的用法让这些伟大文学家们的组合带有一种奇妙的违和感，本应该凸显人物文化底蕴和高雅情操的言辞，反而把这些“作家”们拉下了神坛，表现出一种与粗鄙环境融为一体的和谐，寥寥数语的用典使或迂腐、或自视甚高的“知识分子”形象跃然纸上。而不能忽略的是，莫言所使用的这类名人（即人名）比余华更多，且这种差距在地名上更加明显。当然，这也并不是说莫言小说中的人物就能走遍祖国大江南北、五大洲四大洋，这些地名很多时候与之前所探讨的人名类似，起到形容性的作用，或者是角色的一种吹嘘：

例 8 **内蒙古**来的驼蹄，**黑龙江**来的飞龙，**牡丹江**来的熊掌，**长白山**来的鹿鞭，**贵州**来的娃娃鱼，**威海**来的梅花参，**广东**来的鲨鱼翅（摘自《生死疲劳》）

而不论莫言的地名中有多少是这样的形容、又有多少是实际的地理位置，这些名字的存在本身就拓展了读者的想象范围，因此能够肯定，莫言小说的文本空间一定是比余华笔下的更加广阔的。

类似的情况还出现在重叠式的词汇上：

表 5 余华、莫言特色词中重叠式的对比

形式	余华	莫言
一AA	一笔笔、一户户、一支支、一幢幢、一身身、一拳拳、一箱箱、一封封、一叠叠	一股股、一群群、一丝丝、一摊摊、一步步、一道道、一束束、一拨拨、一绺绺、一滴滴、一幕幕、一幅幅、一枚枚、一堆堆、一摞摞、一座座、一桩桩、一批批、一捆捆、一茬茬、一节节、一瓢瓢、一线线、一截截、一对对、一盘盘、一架架、一缸缸、一年年、一穗穗、一窗窗、一刀刀、一篓篓、一柄柄、一蹿蹿、一双双、一面面、一垛垛、一坨坨、一行行、一株株、一蹦蹦、一声声、一层层、一团团、一道道、一簇簇、一步步、一朵朵、一套套
ABB	好端端、气乎乎、胖乎乎、亮晃晃、暖融融、病歪歪、笑呵呵、乐呵呵、美滋滋、笑吟吟、闹哄哄	乱纷纷、阴沉沉、热辣辣、痴呆呆、水汪汪、油光光、麻酥酥、气汹汹、甜丝丝、慢腾腾、气昂昂、疯颠颠、油汪汪、雾蒙蒙、暖洋洋、凉飕飕、眼巴巴、直呆呆、水灵灵、软塌塌、雾腾腾、凉冰冰、胖嘟嘟、闹嚷嚷、颤悠悠、乌溜溜、木呆呆、密匝匝、病恹恹、大咧咧、油腻腻、忽悠悠、响当当、粘乎乎、紧巴巴、肥滚滚、尖溜溜、麻麻亮、贼溜溜、汗津津、傻呵呵、热哄哄、呆愣愣、直勾勾、胖墩墩、矮墩墩、烂乎乎、清悠悠、平坦坦、稀拉拉、亮光光、颤抖抖、光滑滑、顺溜溜、愣怔怔、笑哈哈、痒酥酥、活泼泼、气鼓鼓、圆乎乎、喜洋洋、气哼哼、好生生、假惺惺、蓬松松、圆溜溜、扑簌簌、急火火、怒冲冲、亮晶晶、明晃晃、热乎乎、辣乎乎
AABB	高高兴兴、亲亲热热、红红火火、空空荡荡、沸沸扬扬、偷偷摸摸、进进出出、支支吾吾、修修补补、指指点点、兢兢业业、认认真真、舒舒服服、漂漂亮亮	轰轰烈烈、影影绰绰、形形色色、原原本本、字字句句、踉踉跄跄、淅淅沥沥、别别扭扭、颤颤悠悠、粘粘糊糊、扭扭捏捏、慌慌张张、密密匝匝、空空洞洞、上上下下、端端正正、重重叠叠、零零碎碎、拉拉杂杂、嘈嘈杂杂、迷迷瞪瞪、丝丝缕缕、咋咋呼呼、疙疙瘩瘩、痴痴呆呆、愣头愣脑、昏昏迷迷、坦坦荡荡、摸摸索索、闹闹哄哄、从从容容、凄凄惨惨、蓬蓬勃勃、平平静静、零零星星、勤勤恳恳、踏踏实实、平平整整、清清爽爽、勉勉强强、迟迟疑疑、普普通通、体体面面、堂堂皇皇、袅袅娜娜、踢踢沓沓、匆匆忙忙、飘飘扬扬、拉拉扯扯、反反复复

不论是“一 AA”式的量词，还是 ABB 和 AABB 式的形容词和动词，重叠式相比原型有加深程度、加强语气、增加感情色彩等表达效果，同时还添加了音韵感。莫言明显多于余华的重叠式类符表现出他文本中更加纷繁的物品、更加鲜明的形态和更加强烈的情感——这正是下一部分的内容。

6.2 莫言强烈的情感和余华克制的情感

以词汇所带有的情绪为标准，挑选余华和莫言词表中的形容词和动词。

表 6 余华、莫言特色词中有感情色彩的形容词、动词对比

类别	余华	莫言
正面动作 动词	弘扬、热恋、凯旋、羡慕、疼爱、称赞、热衷、憧憬	颂扬、歌颂、赞颂、赞同、赞誉、赞许、赞赏、犒赏、艳羡、拥戴、抚慰、眷恋、敬佩、欢呼、感恩、呵护、哺育、留恋
正面情绪 形容词	快乐、高涨	欣喜、愉悦、欢喜、狂喜、欢娱

正面形象 形容词	奢华、柔顺、水灵、细嫩、火爆、热切、勇敢、婀娜、轩昂、壮观、恢宏、庄重	丰富、丰厚、丰腴、丰饶、富饶、富裕、优裕、高尚、高雅、华贵、高贵、尊贵、珍贵、娇贵、娇美、秀美、华丽、秀丽、艳丽、绚丽、炫目、明丽、俏丽、俊俏、俊美、俊秀、灵秀、甜美、可爱、纯正、纯净、香醇、淳厚、淳朴、拙朴、浑厚、厚重、沉静、沉稳、深切、深刻、深邃、温馨、温柔、柔情、柔嫩、柔媚、妖媚、妩媚、性感、娇艳、婉转、凄美、幽婉、凄婉、雄壮、雄伟、宏大、浩大、堂皇、壮烈、恢弘、昂扬、空灵、淡雅、清亮、清朗、明澈、神圣、庄严、圣洁、嘹亮、悠扬、动听、肃静、肃穆、静穆、优秀、优美、高超、绝妙、绝世、豪爽、爽朗、英武、矫健、柔韧、灵动、灵巧、轻巧、机灵、玲珑、细腻、博学、渊博、慈悲、仁慈、鲜亮、热忱
负面动作 动词	摧残、谩骂、危害、糟踏、炸毁、抄家、忏悔、诬蔑、打砸抢	怒吼、怒斥、怒视、嚎哭、嚎叫、叫嚣、悲泣、哀叫、哀嚎、悲鸣、哀鸣、坑害、谋害、侵害、恼恨、憎恨、怨恨、仇视、违规、犯法、违犯、背弃、背叛、残废、残破、残缺、彷徨、失意、污蔑、鄙夷、嘲讽、诽谤、羞辱、蹂躏、践踏、逞凶、作恶、虐待、肆虐、亵渎、催逼、逼迫、威逼、压迫、胁迫、堕落、撕裂、粉碎、杀人、杀害、杀光、枪杀、毒杀、刺杀、绞杀、厮杀、逼死、惨死、屠戮、扫射、毙命、致死、伤亡
负面情绪 形容词	伤心、难受、孤单、悲怆、厌倦、害怕	惊骇、惊慌、惊惧、仓惶、惶惑、惶惶、疑惧、暴怒、愤怒、愤慨、悲愤、困苦、痛苦、哀痛、愁苦、凄苦、哀愁、悲愁、悲凄、悲壮、悲恸
负面形象 形容词	惨重、孤苦、残酷、动荡、贫穷、脏乱、荒芜	狰狞、残酷、暴戾、暴烈、残暴、狂暴、狂躁、暴躁、猖狂、嚣张、狂妄、癫狂、狂乱、遭殃、骚乱、凶残、凶恶、险恶、阴毒、歹毒、拙劣、低劣、恶劣、卑劣、卑鄙、卑贱、邪恶、奸邪、奸诈、诡诈、幽暗、黯淡、浑浊、沉重、低沉、枯涩、颓败、衰败、衰竭、衰退、心碎、恶心、腥臭、腥臊、丑恶、佝偻、肮脏、邋遢、尖酸、龌龊、野蛮、焦躁、绝望、贪婪、虚伪、虚假、荒诞、荒谬、恐怖

表 6 中直观的数量对比当然也表现出莫言丰富的词汇选择空间，但因为其自身带有强烈的情绪——反过来，莫言更多的词汇用量也加强了这种情绪——因此单列一部分专门说明。首先注意到，莫言大量的使用近义的形容词和动词，如表示正面形象中的“丰富”、“丰厚”、“丰腴”、“丰饶”、“富饶”，负面动作中的“哀叫”、“哀嚎”、“悲鸣”、“哀鸣”等等，与表 3 中同音的拟声词情况类似。于是，余华的用词相对则克制很多，例如“快乐”一词出现在他的特色词中，但莫言使用了意义相似但程度不一的“欣喜”、“愉悦”、“欢娱”、“欢喜”、“狂喜”。而若将其中的任何一个换成“快乐”，虽然仍表达喜悦之情，但读者的感受却完全不一样了。

例 9　他**欣喜**地听着锅里的水声，脸上绽开婴孩般的纯洁笑容。（摘自《红高粱家族》）

例 9-改　他**快乐**地听着锅里的水声，脸上绽开婴孩般的纯洁笑容。

在例 9 中，“快乐”显得单纯，“欣喜”却让人感受到此人的内敛性格，莫言就这样用与“快乐”相近但不同的词汇表达了更加丰富的内涵。在莫言的文本中，这样的例子比比皆是，因此读者能够感受到层次更加丰富、程度更加激烈的情感波澜，看到更加细腻、鲜明的文本世界：更多的爱，更多的恨，更多的死亡；美丽的更加美丽，丑陋的更加丑陋——总之，我们也能够看到，莫言文本中所描绘的故事比余华更加浓墨重彩，这也是前人的研究里将莫言的作品形容为“浓墨重彩”的原因吧。

此外，还有一类表示颜色的词汇，兼具了 5.1 丰富的种类和 5.2 强烈的情感，将在第 6 部分集中介绍。

6.3 莫言的文言化和乡土化与余华的白话

余华和莫言的特色词中有一些近义词对尤其能够凸显两位作者的风格差异。

通常认为，文言文向白话文的转变是单音节词双音节化的过程，那么音节数更多的余华——如“或者”对比“或”、“总的来说”对比“总之”，“这时候”对比“这时”，都表现出了更加白话的倾向，相应的，莫言的用词就表现出文言化的特征。除此以外，还有一些典型词汇能够

代表两者的这种差异，一并列入表 7。

表 7 莫言的文言化与余华的白话对比

词类	余华	莫言	词类	余华	莫言
介词	和	与[15]243	助词	似	般[14]2192
连词	和	与[15]243	时间词	今天	今日[18]190，[18]195
	或者	或、抑或[15] 639		中午	正午[20]513
		尔后[15]174			辰巳、乙未、戊寅
	总的来说	总之	时量词	天	日[19]513、宿
副词	越、越来越	愈、愈来愈[15]671	代词	这时候	这时
		动辄[16]555、次第[16]423			吾、汝
		业已[15]559			焉[16] 212，怎[16]240
		曾几何时[17]688			何如[16]1891

词汇上的角标注释了词典中对该词辨析的页码。除了代词中的“吾”、“汝”，时间词中的干支纪年法和量词中的“宿”这种明显的文言词，其他词汇的文言化、书面化特征均能在词典中找到证据。如“似”和“般”这一组近义词，最先出现“一样、相同”这个义项的是“般”：元曲中出现了“高祖般性宽洪，文帝般心圣明”的用法[14]2192，“一般”也出现在了唐代诗歌“湖草青青三两家，门前桃杏一般花”中[16]6；而“似”最早只有“似的”能够表示同样的含义，出现在清朝的《官场现形记》中[16]237。因此可以认为，“般”比“似”更具文言性。

再来看例句：

例 10　到了**中午**的时候，他们终于把屋子收拾干净了（摘自《兄弟》）

例 11　天近**正午**，日头毒辣（摘自《红高粱家族》）

“到了中午的时候”和“天近正午”，是同一意思的不同表达。不仅是在“中午”和“正午”的词汇层面上，还能看到莫言在这一分句上的用词更少、音节更紧凑，这正是文言化的特征之一。类似的情况还出现在副词性语素（dg）上，这是一类修饰谓词性成分（形容词或动词）的语素，其功能与副词相同但不是独立的词汇，因此被命名为副词性语素。例如“狂叫”的“狂”、“尸体横陈”的“横”、“云团上俱镶着灼目金边”的“俱”等等。图 6 中，莫言在这一词类上表现出了显著性，虽然实际的使用频率不高、修饰的对象也主要是动词，但由于莫言和余华在动词的使用频率上无显著差异，可以认为在一些动词的使用上，前者附加了副词性语素而后者没有，这种以语素而非副词的形式进行修饰尽可能少的减少音节、增添了单位词汇的信息含量，表意更加饱满，有着文言化的倾向。

莫言用词的另一特点就是他的方言化、乡土化，例如他比余华使用了更多的后缀“儿”，以及相比余华的“那里”，莫言使用了更多的“那儿”，而儿化音是北方方言的典型特征——在本文开头就介绍过，莫言出生于山东省而余华出身在浙江，生活环境的差异在这一类词汇上体现的非常明显。

表 8 莫言的乡土化与余华的白话对比

词类	余华	莫言
后缀		儿
代词	那里	那儿 这会儿、那会儿 咋[21]1176 俺[21]369
量词		袋子，股子[22]

此外，“咋”、“俺”也是典型的方言词汇，而量词中的“袋子”“股子”这样的双音节词看似是一种白话倾向、与上文所提到的文言化风格相反，但在体词后附加“子”的词汇也具有方言化特征[22]，我们能从中读出人物言行中富含的乡土气息。词汇的对比详见表 8。

需要说明的是，本章不以词类而以作者的特色来进行组织，是为了在更好地突出作者风格的同时不显凌乱。例如，形容词和动词等谓词具有强烈的情感，而以情感为标准来对名词进行分类并不合适，并且不同特征的名词——或其他体词，如代词、时间词等——也体现了作者的不同风格如文言化、方言化等，按词类分节可能导致行文的杂乱，因此组织成本文的模式。

7 颜色词

在检索两位作者的差异词汇时发现了很多颜色词，因此将其单列一节进行讨论。本文选择红、橙、黄、绿、青、蓝、紫、黑、白、灰、金、银、粉、棕共 14 种颜色进行层次聚类，发现颜色词的整体频率的确能够作为区分作者的特征量。

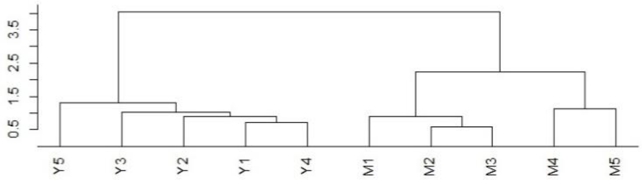
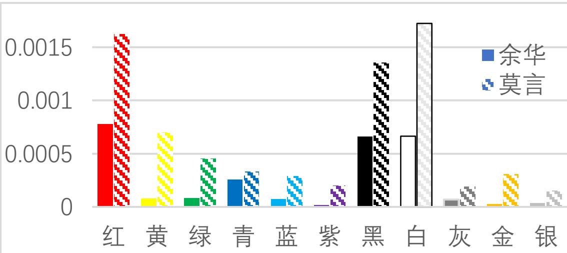


图 8 语料库中颜色频率的层次聚类

两人的颜色词频率对比见图 9。对每一个颜色词都进行构词情况分析，除了自身频率较少的棕色系、橙色系和粉色及构词种类较少可以忽略的紫、银等外，得到的数据见图 10。



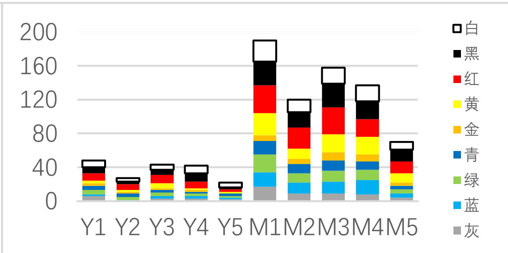


图 9 余华、莫言的颜色词使用频率对比　　图 10 语料库的构词方式种类数量

可以看到，不论是总频率还是构词种类，在每一个颜色词上均有莫言大于余华的情况。而在 6.2 部分也提到，颜色词一类兼具了丰富的类符和强烈的情感两项特征。在此限于篇幅原因，仅以“蓝”为例分析两位作者的不同使用数量、构词和搭配情况，其中“蓝”的数量和构词见表 9。

除“浅蓝”、“深蓝”、“蔚蓝”被余华使用外，其余“蓝”的词汇均仅有莫言使用，且“青蓝蓝”是莫言的自造词。

表 9 10 部小说中“蓝”的构词情况

		Y1	Y2	Y3	Y4	Y5	M1	M2	M3	M4	M5	合
	蓝	12	5	1	5	22	29	27	98	76	17	280
形	浅蓝			2				1	1	7	2	13
容	淡蓝						1		2		1	4
+	深蓝			2	6		4	2		6		20
	纯蓝								1	1		2
	幽蓝						1	1	2	4		8
	暗蓝						1					1
	湛蓝						1	1	11	1	1	15
	蔚蓝	1			1	1	2	6		2		13
实	宝蓝						2			1		3
体	钢蓝						3	1	2	2		8
+	瓦蓝						5	1	3	1		10
	天蓝						4		2			6

	海蓝							1				1
颜	蓝蓝						1		1	1		3
色	青蓝						4		2	2		8
+	灰蓝						6	1	3		1	11
+颜色	蓝灰								1			1
	蓝靛							1	1			2
ABB	青蓝蓝						1					1
	蓝汪汪						2			1		3
	蓝悠悠									1		1
	蓝幽幽							1		1		2
	蓝莹莹							1		1		2
AABB	蓝蓝绿绿						1					1
ABAB	黑蓝黑蓝									1		1
构词方式种类		2	1	3	3	2	17	13	14	17	5	27

从莫言比余华更多的“蓝”的词汇选用上，我们能看到莫言眼中比余华色彩更加丰富的世界。例如同样是形容天空，余华只用来“蓝”和“蔚蓝”，莫言又使用了“蓝蓝”“淡蓝”“湛蓝”“瓦蓝”“宝蓝”“灰蓝”“青蓝蓝”等，而除了同样常被用来形容天空、本就是纯粹的蓝色的“淡蓝”“湛蓝”“瓦蓝”，“蓝蓝”和“青蓝蓝”用重叠的形式加深了蓝的程度，“宝蓝”被用来描述夜晚的天空，“灰蓝”则代表着一头将死的驴子眼中的天空。因此，莫言比起余华更加强调天空的颜色——从所有颜色词中莫言更多的用例来看，可以说是更强调所有事物的颜色——并且会去区分不同天空的蓝，将小说中的世界描摹的更加细致，带给读者更加真实的体验。

莫言还喜欢赋予颜色特殊的情感。他常形容动物的眼睛为蓝色，如：

例 12　这是两只清澈如水的**牛眼，黑蓝黑蓝**的，在漆黑的瞳孔里，我看到了自己的倒影。(摘自《生死疲劳》)

显示出它们不同于人的纯真与无辜。更进一步，莫言还直接将无形的情绪形容为蓝色，这种描述在他的小说中多次出现，它代表着焦虑、愤怒、情欲、痛苦、恐惧种种激烈的情绪，不拘泥于一种。

例 13　上官来弟感到，有一道**蓝色的电火**，从脚踝骨那儿，飞蹿着爬升，并在瞬息间流遍了全身，她全身的皮肤都绷紧了，绷紧了，突然又堤坝决口般地松弛下来。(摘自《丰乳肥臀》)

当这些情绪在脑海中骤然爆发出来，莫言赋予了它们幻觉般的蓝色，让它们出现在角色的心中，也出现在读者的眼前。其他未在文中列出的 13 种颜色，表现也与“蓝”类似。

莫言丰富的构词为读者展现出他小说世界中层次丰富色彩，超常的又表达出充沛的情感。同时，这些颜色不仅单独出现，还会一起组成丰富的意象群，尤其是对比色的聚集：

例 14　他想起那个**翠绿**的高粱地里的**火红**的中午。想起那匹陷在窗外泥土里的**黑**骡子，想起**白净**的肉体躺在黑眼的怀抱里。(摘自《红高粱家族》)

这些鲜明对比的颜色给人强烈的视觉冲击感，让整幅画面更加生动鲜活，不论是画面本身还是其背后蕴含的情绪，在给人带来超现实的魔幻体验的同时隐藏着作者渲染情绪的意图，给整个场景染上或悲壮或污浊或艳情或恐惧、种种不同的感觉。

8 结语

本文出了比较两位作者在段落、句子、分句和标点上的使用差异外，主要工作集中在词汇上。首先通过假设检验的方式对余华和莫言的词表进行分类，再通过寻找近义词对的方式比较两人的语言风格倾向，发现相比余华，莫言的用词表现出了文言化、方言化的特征，有更加丰富的词汇选择空间、更加充沛和细腻的情感表达和形象描摹，因此在文本中包含了更加丰富的信息量。另外，还选择颜色这一语义场作为对象，研究了两位作者不同构词和搭配的情况，发现了莫言在颜色词上更加多样的构词、更多的超常搭配和连用情况，及以此构筑

的超现实文本世界和带给读者的强烈感官刺激。

因此，相比于莫言个人特色的浓烈，余华的文本风格则相对更加白话、冷静一些。一方面，正是这种冷静使两人同样具有的戏谑风格凸显出讽刺的感觉；另一方面，这也是文本风格学研究的重要特征：风格只有在对比中才能体现。因此，对本文所提方言化、文言化等特征，一方面很难得穷尽地找到它们背后的代表词汇，一方面其所描绘的作者风格还是相对的、必须要有其他文本对照才能确认的。

最后，本文所使用的文本聚类属于无监督的学习方法，限于篇幅无法再运用有监督的文本分类进行风格预测工作——这将是我们下一步的研究方向。

参考文献

[1] 徐文明. 死亡的风景——余华、莫言暴力叙述现象研究[D]. 河南大学，2005.

[2] 李艳君. 悲凉与悲壮的人生画卷——简评余华《活着》与莫言《丰乳肥臀》[J].青年文学家，2014，29：12-13.

[3] 毛冰. 莫言与余华小说阅读感受的异同[J].文学教育，2014，5：94-96.

[4] 王清溪，付维洁. 先锋作家余华小说语言新变探析[J].牡丹江教育学院学报，2006，6：19-20.

[5] 王委艳. 论余华小说的陌生化叙事[J].康定民族师范高等专科学校学报，2006,15（2）：45-48.

[6] 季红真. 现代人的民族民间神话——莫言散论之二[J]. 当代作家评论，1998，4：80-89.

[7] 杨连芬. 莫言小说的价值与缺陷[J].北京师范大学学报，1990，1：63-69.

[8] 刘颖，肖天久. 金庸与古龙小说计量风格学研究[J].清华大学学报(哲学社会科学版),2014,29(05):135-147+179.

[9] Kerner, Y. H., Margaliot, O. AUTHORSHIP ATTRIBUTION OF RESPONSA USING CLUSTERING[J]. Cybernetics & Systems, 2014, 45(6):530-545.

[10] 贺湘情，刘颖. 基于文本聚类的语言韵律和节奏风格特征挖掘[J]. 中文信息学报、2014，28(6):194-200.

[11] Sapkota, U., Bethard, S., Montes, M., et al. Not All Character N-grams Are Created Equal: A Study in Authorship Attribution[C]. Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2015.

[12] 吴喜之. 非参数统计[M]. 北京：中国统计出版社，2006.

[13] Grieve, J. Quantitative Authorship Attribution: An Evaluation of Techniques[J]. Literary and Linguistic Computing, 2007, 22(3): 251-270.

[14] 阮智富,郭忠新 编著.现代汉语大词典·下册[M].上海：上海辞书出版社. 2009.

[15] 张斌. 现代汉语虚词词典[M]. 商务印书馆，2005.

[16] 徐复等 编.古代汉语大词典[M].上海：上海辞书出版社.2007.

[17] 郝迟,盛广智,李勉东 主编.汉语倒排词典[M].哈尔滨：黑龙江人民出版社.1987.

[18] 张清源 主编.现代汉语常用词词典[M].成都：四川人民出版社.1992.

[19] 汉语大词典编纂处. 汉语大词典(普及本)[M]. 上海辞书出版社. 2000.

[20] 王海棻.古汉语时间范畴词典[M]. 合肥：安徽教育出版社.2005.

[21] 阮智富,郭忠新 编著.现代汉语大词典·上册[M].上海：上海辞书出版社. 2009

[22] 昌梅香.现代汉语方言后缀研究[J].学术交流,2008(01):132-135.

作者联系方式：

涂梦纯 北京市海淀区清华大学中文系 100084 18810302620 toodreamyclean@163.com

刘 颖 北京市海淀区清华大学中文系 100084 18500430166 yingliu@tsinghua.edu.cn