

海明威和菲茨杰拉德小说词汇计量风格学的对比研究

罗茵林^{*} 韦爱云

广西师范大学 外国语学院, 中国·广西 桂林 541006

摘要: 词汇的计量特征研究一直以来都是计量风格学的研究重点之一。基于自建 30 多万词的海明威和菲茨杰拉德小说语料库, 本研究一方面考察两位作家小说在词汇层面的总体计量特征, 结果表明海明威在词汇使用方面比较单一、简练; 菲茨杰拉德则更善于使用华丽的词藻。另一方面, 分析了两位作家在具体词汇如主题词、颜色词及情态动词使用上的异同。结果表明海明威更多的使用酒精去展示战后一代的迷惘; 菲茨杰拉德则更注重神态、声音的描写, 且更擅长用大量的颜色描写丰富文章的象征主义; 在情态动词的使用方面, 海明威倾向于使用中量值情态动词, 避免语气比较强硬; 而菲茨杰拉德则更多使用低量值情态动词, 表达上更显委婉。

关键词: 海明威; 菲茨杰拉德; 词汇; 计量风格

A Comparative Research of Lexicometric Stylistics in Hemingway and Fitzgerald's Novels

Yinlin Luo^{*} Aiyun Wei

Guangxi Normal University, College of Foreign Studies, Guilin, Guangxi, 541006, China

Abstract: The study of the quantitative features of vocabulary has always been one of the research focuses of quantitative stylistics. Based on a self built corpus of over 300000 words of Hemingway and Fitzgerald novels, this study examines the overall quantitative characteristics of the two authors' novels at the lexical level. The results show that Hemingway's vocabulary usage is relatively simple and concise; Fitzgerald was better at using fancy language. On the other hand, the similarities and differences in the use of specific vocabulary such as theme words, color words, and modal verbs between the two writers were analyzed. The results indicate that Hemingway used more alcohol to showcase the confusion of the post-war generation; Fitzgerald, on the other hand, pays more attention to the description of demeanor and voice, and is better at using a large number of colors to enrich the symbolism of the article; In terms of the use of modal verbs, Hemingway tends to use medium value modal verbs to avoid a more forceful tone; Fitzgerald, on the other hand, uses more low value modal verbs and expresses himself more tactfully.

Keywords: Hemingway; Fitzgerald; vocabulary; measurement style

1 引言

海明威和菲茨杰拉德是“迷惘的一代”的重要代表, 他们的创作主题和写作风格各具特色, 但都反映了那个时代的迷惘和彷徨。国内外已有许多学者对比他们在创作上的异同: 有针对两位作家主题创作方面的比较研究^[1,2]; 有从女性观角度对比两位作家作品的研究^[3]; 有从语言使用的角度对比两位作家作品的定性研究^[4]。还有部分学者尝试从量化的角度对两位作家作品进行比较研究^[5,6]。综上, 大部分针对海明威和菲茨杰拉德的比较研究是定性的, 虽然也有学者尝试从定量的角度进行比较, 但从词汇计量特征角度对两人作品的语言风格进行系统对比研究的还比较罕见。因此, 论文打算从词汇层面对两者的小说进行量化研究。探讨以下研究问题:

①海明威和菲茨杰拉德的小说在词汇层面的总体计量特征上有何异同?

②海明威和菲茨杰拉德在具体词汇如主题词和颜色词等使用上有何异同?

2 语料与方法

2.1 语料来源

论文语料来源于 Z-library。在选取语料时, 考虑到文本长度及作品发表年份可能会影响定量分析的结果, 因此选择海明威的《太阳照常升起》和《永别了, 武器》以及菲茨杰拉德的《夜色温柔》和《了不起的盖茨比》。选取的语料长度相当, 发表年份都在 20 世纪 20—30 年代, 且在学界上都引起了学者的广泛关注。表 1 是两个语料库的基本统计信息。

表 1 海明威库和菲茨杰拉德库基本统计信息

	形符 (token)	类符 (type)	类符形符比 (type/token ratio)
海明威库	157219	7510	4.78%
菲茨杰拉德库	161636	13158	8.14%

2.2 研究方法

本研究基于两个自建的小型语料库, 从词汇层面对两个语料库进行量化对比研究, 分析两者的异同及特点。研究

分为两个步骤:

第一步,收集相关语料后,我们对语料进行了人工处理、校对,删去特殊符号、空格及与正文无关的内容后,根据研究内容创建了语料库,分别为海明威库和非茨杰拉德库。为了解两位作家的小说在词汇层面的总体计量特征上的差异,我们使用 WordSmith 生成文本的平均词长、标准类/形符比、单现词;使用 Readability Analyzer 2.0 计算文章的可读性;并用 python 自写程序爬取文章的词汇密度,对两位作家的小说作品总体的词汇计量特征进行综合的分析。

第二步,为了比较两者在具体的词汇如主题词、颜色词和情态动词的应用方面可能存在的差异,我们使用 python 抓取频次排名前 20 的主题词,并爬取文本中颜色词及情态动词出现的频次,结合相关的计量语言学理论进行分析。

3 结果与讨论

3.1 词汇层面总体的计量特征分析

通过统计,两个语料库的标准类/形符比、平均词长、可读性、单现词的数量如表 2 所示。

表 2 两个语料库的标准类/形符比、平均词长、可读性、单现词数量表

语料库	标准类/形符比	平均词长	可读性	单现词	单现词比例
海明威库	35.49%	3.96	2.89	3213	2.04%
菲茨杰拉德库	46.14%	4.25	7.09	6443	3.97%

3.1.1 标准类/形符比

类符是文本中不重复计算的词,形符指的是文本中的所有词形。类/形符比的高低与文本词汇丰富度成正比关系。即类/形符比越大,文本词汇量越丰富^[7]。然而由于语料库的容量不同,类/形符比不一定能说明问题,故论文采用 Scott^[8]提出的标准类/形符比,即每千字类符/形符比的均值。海明威库的标准类/形符比 35.49% 远远小于菲茨杰拉德库的 46.14%,即海明威使用的词汇比较单一。刘泽权、王梦瑶^[9]分析海明威作品《老人与海》的词汇特征也得出过海明威作品词汇变化小的特点。而菲茨杰拉德所使用的词汇相对来说非常丰富,这主要是因为他的创作风格相对来说更精致和优美,使用的词汇变化多样,从而使得作品更加生动和具有感染力。

3.1.2 平均词长

平均词长是文本中所有单词的平均长度。海明威库的平均词长为 3.96,而菲茨杰拉德库的平均词长为 4.25。一般来说,平均词长较长则表示该文本长词较多,理解起来比较复杂;相反,如果词汇长度较小,则比较容易理解^[10]。结果表明海明威用词比较简短,这与汪琴、刘富丽^[11]基于海明威四部长篇小说的量化研究结果一致,说明海明威的作品比较容易理解;而菲茨杰拉德使用的词汇较长,相对来说较

难理解。

3.1.3 可读性

作品的可读性用 Readability Analyzer 2.0 测出,并用其中的 Flesch Kincaid Grade Level 指标^[12]来表示。该指标通过计算每个句子的平均单词数和单词的平均音节数来反映文本的复杂程度。具体来说,反映文本复杂程度的分数与文本的复杂程度成正比关系。分数越高,读者阅读难度就越大。海明威库的可读性为 2.89,远远小于菲茨杰拉德库的 7.09,可见菲茨杰拉德作品的阅读难度比海明威的要高,且难度是他的两倍多。

3.1.4 单现词

单现词即文本中只出现了一次的词,是另一个可以衡量文本词汇丰富度的参数^[13]。单现词与文本词汇密度呈正比关系,单现词越多,证明文本词汇越丰富。海明威库的单现词有 3213 个,比例为 2.04%,而非茨杰拉德库的单现词有 6443 个,比例为 3.97%。菲茨杰拉德库的单现词比例大于海明威的,这主要是因为海明威受到一战期间军事经历和记者背景的影响,使他在写作中形成了简洁的“电报式”风格。因此在他的作品中,无论是写人、描绘景物,还是描写人物对话,都采用了新闻性凝练的语言^[14],所使用的词汇常见普通,以避免过度修饰。结果与康璐、钟晓红^[15]以及刘艳^[16]等学者采用定性方法对海明威小说语言风格的研究结果一致,说明这是海明威的作品中普遍存在的特点。而菲茨杰拉德的作品中常常存在着情感与理性的冲突、现实与幻想的对立、战后一代青年的梦想与失意等主题,主题丰富,涉及的领域也多。且光明与黑暗、美丑善恶之间的激烈冲突及反差都有呈现,发挥出了他所特有的创作风格和叙事艺术^[17],故词汇也更加丰富。

3.1.5 词汇密度

词汇密度可以用于衡量文本信息量的大小。词又分为实词和虚词,理论上来说,一个文本的实词越多,其涵盖的语义信息就越复杂,提供的文本信息也越多。实词由动词、名词、形容词、副词构成。词汇密度的计算有多种方式。论文采取的是 Stubbs^[18]提出的方式,即只计算实词与总形符的比,具体计算公式如下:

$$\text{词汇密度} = \frac{\text{实词数}}{\text{总形符数}}$$

通过观察表 3 可知,菲茨杰拉德库的词汇密度(42.83%)比海明威库(42.48%)的要高,说明菲茨杰拉德作品的实词较多,所提供的信息也更加丰富。且通过观察两位作家实词词类的比例,可以发现他们在词汇使用方面的风格。菲茨杰拉德在名词的使用上比海明威要多,这主要是因为他以华丽的词藻和丰富的象征主义而闻名,擅长用大量的名词去描写人物、场景等。而海明威的创作风格则以简洁著称,在他的作品中,名词的使用通常更加朴素和直接,力求简洁明了。

形容词主要是对名词进行修饰，因此形容词与名词的比例是成正比关系^[19]，故菲茨杰拉德的形容词使用比例（14.06%）

也多于海明威（13.53%）的。同时他还善于使用各种修辞手法，这都使得作品中的语言变得更加丰富、生动、形象^[20]。

表 3 海明威库和菲茨杰拉德库词汇密度对比表

词类	海明威库	菲茨杰拉德库	海明威库词汇密度	菲茨杰拉德库词汇密度
名词	35.72%	38.01%		
动词	12.55%	15.16%		
形容词	13.53%	14.06%	42.48%	42.83%
副词	38.20%	32.77%		

3.2 具体词汇如主题词和颜色词等使用分析

虽然海明威和菲茨杰拉德处于同一个时代，但他们的写作主题却各具特色，写作主题可以通过主题词加以分析；而颜色则富有象征主义，颜色对于揭示主题、对故事进行烘托渲染具有重要作用；此外，情态动词在文学作品中可以表达人物角色的立场态度、还可以传递情感。综上，将分析两位作家在主题词、颜色词和情态动词的使用上可能存在的异同，以更全面的揭示两位作家的创作风格及写作特点。

3.2.1 主题词

不同的文本在词语的选择上有所差异，主题词则可以通过作者对词语的选择而表现出来^[21]。主题词一般来说是名词，故提取两位作家作品频次前 20 的名词（人物名字除外）分析他们在主题方面的异同。

表 4 海明威库和菲茨杰拉德库频次前 20 的名词

语料库	单词频次前 20 的名词
海明威库	time; room; road; bull; night; bed; people; way; darling; side; door; man; wine; table; street; head; town; hand; war; face
菲茨杰拉德库	man; time; people; eyes; moment; room; face; way; night; girl; house; men; voice; car; door; doctor; years; day; hand; thing

首先通过观察表 4 可知，两者都有“night”这一主题词，说明他们的故事发生时间也许多在夜晚。其次，海明威使用较多的词有“road/room/street”，说明故事在室内室外都有发生；“wine”一词则表示海明威大量的使用酒精去展示战后迷惘一代的颓废、享乐。而菲茨杰拉德则大量的使用了“room/house”，表示故事以发生在室内居多，且文中大量奢华的场景都是在室内，具有重要的象征意义。最后，菲茨杰拉德更多的使用“face/voice/eyes”，说明他可能更注重神态、声音的描写。陈喜华^[22]分析《了不起的盖茨比》中黛西形象建构的定性研究也曾得出类似的结论，在《了不起的盖茨比中》，菲茨杰拉德在黛西的声音、神态描写方面下了不少功夫。虚伪的女低音揭示了黛西索取感情、内心充满金钱的本质；在神态方面，黛西特有的表情做作且不自然，展现出她放荡不羁、自私冷漠的特点。通过神态和声音的描写，凸显出菲茨杰拉德对人物性格的精准把握。

3.2.2 颜色词

象征主义经常会被用于各种文学作品中。菲茨杰拉德

和海明威的作品就出现了颜色这一象征符号，通过统计 13 种常见颜色：红、橙、黄、绿、蓝、紫、黑、白、灰、金、银、粉、棕在两人作品中出现的频率，制作出两人作品的颜色词频率对比图。

颜色词对比条形图

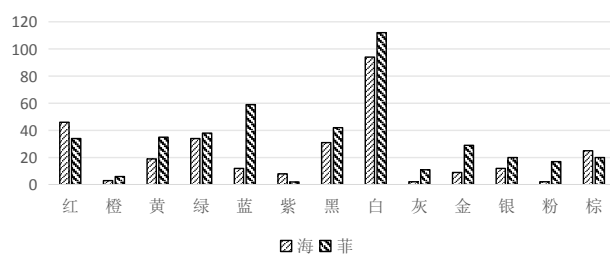


图 1 海明威库和菲茨杰拉德库颜色频率对比图

由图 1 可以看出，在 13 种常见颜色中，菲茨杰拉德在大部分的颜色使用上频次都要高于海明威，尤其是黄色、蓝色、金色、银色、粉色这些亮色。菲茨杰拉德的写作风格更为华丽，富有象征主义，能够给读者一种眩目的感觉，突出上流社会和富人阶层奢靡享乐、纸醉金迷的生活。在《了不起的盖茨比》中“His gorgeous pink rag of a suit made a bright spot of colour against the white steps”以及“*And Gatsby, in a white flannel suit, silver shirt, and gold-coloured tie*”^[23]可以看出菲茨杰拉德在服饰描写方面使用颜色非常活泼、华丽，以突出人物的社会地位。

而海明威则在红色以及紫色两种颜色的使用上频次明显要高于菲茨杰拉德。红色作为一种鲜艳的亮色在海明威作品中占比较高的原因是：海明威一生中大部分的时间都在喝酒，无论是借酒消愁，还是开怀畅饮，他在作品中经常以自己的经历为素材。且酒对于人物形象的塑造也起着一定的作用，从海明威大量描写喝酒的场景就可以窥探出主人公的内心世界，因此作品中会提及各种各样的酒^[24]，“red wine”就是其中一种，即葡萄酒。海明威生动的刻画了迷惘一代将酒当作麻醉剂的这一特征，突出战后一代身心受到深远影响。如《太阳照常升起》中，主人公杰克在一战中受了伤，在身体和心理上都茫然失落，每天漫无目的的在巴黎的各个酒吧喝的酩酊大醉，利用酒精排解心中的烦恼，被斯泰因称为“迷惘一代”的典型代表^[25]。而紫色频次比较多的原因是海明威在《太阳照常升起》中多次提到威廉·亨利·赫德

逊的“The Purple Land”一书，对于没有参加过一战的科恩来说，这本书燃起了他探索南美的热情。海明威成功刻画了一个有理想，却只是空想；有浪漫情怀，却很传统陈腐的形象，这也属于迷惘的一代青年中的其中一类^[26]。

3.2.3 情态动词

情态动词是用来表达情感、态度或者可能性的一类动词，在海明威和菲茨杰拉德的作品中就出现了较多的情态动词。学界已有许多学者对情态动词进行分类^[27,28,29]，Halliday把情态系统分为高、中、低三个量值，高量值的情态动词有 must, ought to, need, have/has/had to；中量值情态动词有 will, would, shall, should；低量值情态动词有 may, might, can, could，划分比较具体，因此本研究主要以Halliday^[29]的情态动词分类为标准，探讨海明威和菲茨杰拉德在情态动词使用上的异同。两个语料库所使用情态动词的具体情况如表5所示。

表5 海明威库和菲茨杰拉德库情态动词频次图

情态量值	情态动词	海明威库	占比 (%)	菲茨杰拉德库	占比 (%)
高	must	124	5.99	125	9.27
	need	29	1.40	15	1.11
	ought to	64	3.09	47	3.49
	Have/has/had to	155	7.49	87	6.45
合计		372	17.97	274	20.32
中	will	367	17.73	104	7.72
	would	516	24.93	297	22.03
	shall	6	0.29	18	1.34
	should	137	6.62	65	4.82
合计		1026	49.57	484	35.91
低	can	189	9.13	195	14.47
	could	397	19.18	286	21.22
	may	22	1.06	31	2.30
	might	64	3.09	78	5.79
合计		672	32.46	590	43.78
总计		2070	100	1348	100

由表5可知，两位作家在情态动词的使用上既有相同点也有不同点。首先，海明威在情态动词的使用上要远多于菲茨杰拉德，分别为2070次和1348次。通过对语料库进行检索可知，情态动词绝大部分出现在人物对话或者人物描写中。情态动词的使用对人物性格的刻画以及角色之间的关系描写起着画龙点睛的作用^[30]，说明海明威在人物刻画上可能更偏向于使用情态动词去突出角色的特点。其次，两者都较少的使用高量值情态动词；海明威更注重中量值情态动词的使用，这在他所使用的三个量值的情态动词中占比为49.57%；菲茨杰拉德则更注重低量值情态动词的使用，占比为43.78%。情态量值表示说话人对听话人压力施加的程度。也就是说，说话人使用高量值的情态动词时，对听话人施加的压力最强；而低量值情态动词表示委婉语气，施加

的压力最小^[31]。这说明在海明威的作品中，描写人物之间的对话或场景时，他会尽量避免使用对听话人施加压力最强的词，更多使用“will, would”这类中量值情态动词，防止语气过于强硬及绝对。如在《永别了，武器》中，主人公Henry从驻扎的小镇回到前线，与他合住的中尉Rinaldi对他在小镇的生活非常感兴趣，因此他对Henry说：“Tonight you will tell me everything.”^[32]此处并没有用强硬的语气去命令Henry，而是用情态动词“will”，既表现了他的迫切，又没有强迫之意。而菲茨杰拉德则更多的使用低量值情态动词，如在《了不起的盖茨比》中，麦基太太想让麦基先生给妹妹凯瑟琳拍照，说：“Chester, I think you could do something with her.”^[33]此处使用了情态动词“could”，这使得说话人表达的意思比较委婉，或透露出含糊的语气^[33]，所发出的指令压力强度小。

4 结语

论文基于自建的海明威和菲茨杰拉德小说语料库，采用量化分析的方法对两位作家小说的词汇层面进行了分析，得到的结论如下：

①菲茨杰拉德小说的词汇丰富度要高于海明威的，且平均词长比较长，阅读难度比海明威的要高很多。

②在主题词方面，两者描写的场景以夜晚居多，海明威的故事在室内室外都有发生，而菲茨杰拉德的故事则发生在室内居多。海明威擅长用酒去凸显迷惘一代的颓废和放纵；菲茨杰拉德则更注重神态、声音的描写。在颜色词方面，菲茨杰拉德使用颜色的频率更高，象征意义丰富，相对来说情节更加生动。在情态动词的使用方面，海明威更倾向于使用中量值情态动词，避免语气过于强硬；而菲茨杰拉德则更多使用低量值情态动词，表达的意思比较委婉。

本研究表明，采用量化的方法比较作家的词汇使用风格更加客观且具有说服力。但是，由于语料规模还不够大，且论文仅侧重于两位作家在词汇方面的对比分析，研究仍存在一定的缺陷。希望将来的研究可以扩大语料规模，增加量化指标，从更多视角去比较两位作家的写作风格。

参考文献：

- [1] 孔耕蓰. 冰山与夜色的对话——海明威和菲兹杰拉德艺术世界透视[J]. 外国文学评论, 1994(3): 84-89.
- [2] 熊文, 秦秋. 共同的迷惘不同的探索——海明威、菲茨杰拉德之比较[J]. 江苏社会科学, 2005(5): 194-198.
- [3] 周彦. 两个女主角 两种妇女观——从黛西和白瑞特看菲茨杰拉德和海明威的妇女观[J]. 广西民族学院学报(哲学社会科学版), 1995(1): 107-109.
- [4] 谷野平. 海明威与菲茨杰拉德的小说创作异同析[J]. 辽宁师范大学学报(社会科学版), 2007(3): 87-89.
- [5] Rice J. What Makes Hemingway Hemingway?[J]. A stational analysis of the Data Behind, 2016.
- [6] Deli Z P. The lexical analysis of two works by Ernest Hemingway

- and F. Scott Fitzgerald[J]. *Porta Lingua*,2021(1):199-209.
- [7] Baker M. Towards a methodology for investigating the style of a literary translator[J]. *Target. International Journal of Translation Studies*,2000,12(2):241-266.
- [8] SCOTT M. *The WordSmith Tools*[M]. Oxford: Oxford University Press, 2004.
- [9] 刘泽权,王梦瑶.多变量方法在文学风格考察中的应用——以《老人与海》为例[J].*外语电化教学*,2017(5):61-67+92.
- [10] 李敏杰,朱薇.基于平行语料库的《红楼梦》英译论文体风格研究[J].*中南民族大学学报(人文社会科学版)*,2012,32(2):177-180.
- [11] 汪琴,刘富丽.海明威长篇小说的语料库文体学分析[J].*台州学院学报*,2011,33(5):37-39+59.
- [12] Kincaid J P, Aagard J A, O'Hara J W, et al. Computer readability editing system[J]. *IEEE Transactions on Professional Communication*, 1981(1):38-42.
- [13] Daller H, Van Hout R, Treffers - Daller J. Lexical richness in the spontaneous speech of bilinguals[J]. *Applied linguistics*, 2003, 24(2):197-222.
- [14] 黄俏.从记者到作家——试论海明威小说的“新闻”性写作[J].*新闻爱好者*,2009(20):188.
- [15] 康璐,钟晓红.论海明威小说语言艺术风格的形成及其特色[J].*语文建设*,2013(33):21-22.
- [16] 刘艳.对海明威小说语言风格的解读[J].*语文建设*,2014(32):57.
- [17] 吴建国.菲茨杰拉德创作简论[J].*上海海运学院学报*,2001(4):93-105.
- [18] Stubbs M. Language development, lexical competence and nuclear vocabulary[J]. *Educational linguistics*,1986,7(1):1-18.
- [19] 沈威,吴玉容.巴金和茅盾小说词汇计量风格学研究[J].*四川文理学院学报*,2021,31(4):63-71.
- [20] 田然.菲茨杰拉德小说语言特色解读[J].*语文建设*,2016(23):63-64.
- [21] 张海云,谢群芳.基于语料库的文学作品检索分析——以越南中篇小说《志飘》为例[J].*解放军外国语学院学报*,2010,33(3):57-61.
- [22] 陈喜华.匠心独妙之笔——《了不起的盖茨比》中的黛西的形象建构[J].*郑州航空工业管理学院学报(社会科学版)*,2012,31(4):38-40.
- [23] Fitzgerald F S. *The Great Gatsby*[M]. Hertfordshire:Wordsworth Editions Press,1999.
- [24] 杨雁,王群锋.被酒充斥的世界——论酒在《永别了,武器》中的作用[J].*贺州学院学报*,2007(1):66-69.
- [25] 邱世凤.论海明威的酒情结[J].*电影评介*,2007(9):94-95.
- [26] 王震宇.《太阳照常升起》的主题内涵与艺术表现[J].*语文建设*,2016(29):69-70.
- [27] Biber D, Johansson S, Leech G, et al. *Longman Grammar of Spoken and Written English*[M]. London: Pearson Education,1999.
- [28] Palmer F. *Mood and Modality*[M]. Cambridge: Cambridge University Press,1986.
- [29] Halliday M A K, Matthiessen C M I M. *An Introduction to Functional Grammar*[M]. 3rd Edition, London:Hodder Arnold,2004.
- [30] 郑窈娜.从人际功能的情态系统角度分析《呼啸山庄》中的人物关系[J].*咸宁学院学报*,2010,30(4):57-58.
- [31] 李鑫,胡开宝.基于语料库的记者招待会汉英口译中情态动词的应用研究[J].*外语电化教学*,2013(3):26-32+74.
- [32] Ernest. Hemingway. *A Farewell to Arms*[M]. New York: Charles Scribner's Sons,1957.
- [33] 张晓燕.语料库文体学视角下《喜福会》的N词语块分析[J].*东北农业大学学报(社会科学版)*,2015,13(5):93-98.
- 作者简介: 罗茵林(2000-),女,中国广西钦州人,硕士,从事计量语言学研究。
- 韦爱云(1976-),女,中国广西东兰人,博士,教授,从事计量语言学、英语教学、翻译、文学研究。