doi:12. 3969/j. issn. 1672-0598. 2012. 02. 021

语料库语言学视角下的文学批评
——以《红字》为个案分析*

高　博

(天津科技大学 外国语学院,天津 300222)

摘　要:语料库语言学作为一门新兴学科,已经被广泛运用于文学批评领域。以 BROWNA 语料库中小说部分为参照,建立《红字》可比语料库,利用语料库检索分析软件 WordSmith Tools 4.0,对两个语料库的平均词长、词性分布、主题词和平均句长等信息进行统计和量化分析。分析结果不仅验证了 Hawthorne 独特的文体风格,同时也揭示了《红字》的故事情节。

关键词:语料库语言学;文学批评;《红字》;霍桑

中图分类号:I06　**文献标志码**:A　**文章编号**:1672 - 0598(2012)02 - 0130 - 06

一、引言

文学作品是人类社会的宝贵财富,它不仅记录了人类历史的发展进程,同时也承载了人们的智慧结晶。然而,在文学作品的传播过程中,越来越多的读者在欣赏文学作品内容的同时,也开始了对它们进行系统的研究和评价并形成了成熟的理论体系,即文学批评理论。具体来讲,文学批评是指按照一定的标准对作家作品和文学现象(包括文学运动、文学思潮和文学流派等)所作的研究、分析、认识和评价。它以文学鉴赏为基础,同时又是文学鉴赏的深化和提高。在文艺学的诸种研究形态中,文学批评是最活跃、最经常、最普遍的一种(蒋原伦、潘凯雄,2006)。但是,传统的文学批评理论通常具有很强的主观性,它常以研究者对于文学作品的直观感受作为主要衡量标准,缺乏客观,翔实的数据作为支撑。因此,Fowler(1975)提出了"新文体学"这一概念,呼吁当代语言学理论及技巧应用到文学的研究中来。在这样的背景之下,基于语料库的文学批评研究逐渐走进了人们的视野。

二、语料库语言学与文学批评

20 世纪 60 年代以来,伴随着计算机技术的发展,在语言研究领域中出现了一种全新的交叉学科,即语料库语言学。"语料库语言学是以真实的语言数据为研究对象,从宏观的角度对大数量的语言事实进行分析,从中寻找语言的规律;在语言分析方面采用概率法,以实际使用中的语言现象的出现概率为依据建立或然语法进行语法分析"(杨惠中,2002)。语料库语言学的出现为文学批评研究提供了崭新的视角。基于语料库的文学批评研究通常利用文学作品建立语料库,使用语料库检索分析软件,以文学语言和文学结构作为研究对象,通过用词分布分析、文本特征分析和情节分析等计算机统计分析技术,拓展传统的文

* [收稿日期]2011 - 12 - 21
[作者简介]高博(1986—),男,天津人;天津科技大学外国语学院硕士研究生,主要从事语言学、应用语言学和语料库语言学研究。

学研究,提炼文学修辞和文学叙事的规则等。除此之外,还可以通过建立可比语料库,将某一特定作家的文学作品与其他作家的文学作品进行对比,进而甄别出该作家独特的文体风格。

近年来,基于语料库的文学批评研究获得了文学研究者越来越多的青睐,与其有关的论文数量也在逐年增长。这些研究主要集中在:(1)文学作品的情节分析,如杨建玫(2002),周艳丽、张发祥(2008);(2)文学作家的文体分析,如马广惠(2005),David. L. H(2010);(3)语料库与文学教学,如张显平(2007);(4)语料库与文学研究综述,如李晋、郎建国(2010)等。笔者借鉴了基于语料库的文学批评方法,以19世纪美国浪漫主义大家Hawthorne的文学巨著《红字》为研究对象,试图探索Hawthorne的文体风格及其形成原因。

三、本研究所使用的语料库及研究方法

本研究以《红字》为研究对象,采取定量与定性相结合的方法,通过建立可比语料库进行实证性研究。该语料库主要包括两个部分,即《红字》语料库以及与其进行对比的文学作品参照语料库。笔者通过网站 http://www.hjenglish.com/dl/dl6007 下载了《红字》的电子文档并建立了《红字》语料库,共计69,221词,同时,为了保证研究的客观性,笔者抽取了BROWNA(普通英语语料库)语料库中的小说部分作为文学作品参照语料库(以下称之为参照库),该语料库含词量为359,993词。本研究使用的语料库检索软件是WordSmith Tools 4.0。此外,根据研究需要,这两个语料库都使用了CLAWS词性附码器和分类详细的CLAWS7词性附码集做了词性附码。

四、语料的分析与讨论

(一)平均词长

平均词长是指特定语料库中形符的平均长度,以字母数量为单位。平均词长越长,说明文本中使用的长词越多,文本的阅读难度相对就越大(陈建生、高博,2011)。表1统计了两个语料库的词长信息。

表1 两个语料库的平均词长

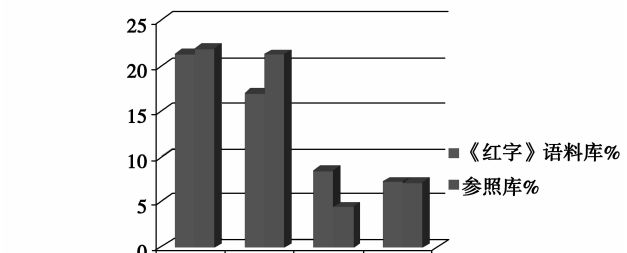
	《红字》语料库	参照库
平均词长	4.51	4.03

由表1可知,《红字》语料库的平均词长为4.51,而参考库的平均词长为4.03,独立样本t检验结果表明,两个语料库在平均词长使用上存在显著性差异($t = 3.32$, $p < 0.01$)。这个结果说明,相对于其他文学作品,Hawthorne的《红字》更加倾向于使用比较复杂的词汇,文体风格更加正式,阅读难度更高。例如:《红字》第十八章写到:"It was the exhilarating effect-upon a prison just escaped from the dungeon of his own heart-of breathing the wild, free atmosphere of an unredeemed, unchristianized, lawless region."(这对一个刚摆脱作茧自缚的人来说,是一种振奋的力量,无异在一个尚未开发、尚未基督教化、尚无法律的境界里呼吸到一阵空旷、自由的空气)。这段话是用来描述青年牧师Dimmesdale在经历过巨大的心理挣扎之后的一种超然的精神解脱。不难看出,Hawthorne在该句之中使用了大量的合成词,尤其是当他形容牧师的超然心境时,更是连续使用了三个合成形容词,即unredeemed,unchristianized和lawless。这三个词的使用凸显了剧中人物与其社会身份以及社会背景的矛盾,Dimmesdale作为一名基督教牧师,理应严肃、虔诚和公正,但是,正是这样的一名牧师却被一种背离了基督教义的力量所感动,所震撼。Hawthorne在描述牧师这种振奋的心情时,没有选择使用背离基督教义和准则的贬义词汇,如heretic, rebellious等,而是选择了在原有基督教准则的词汇(redeemed, christianized and law-abiding)之上直接冠以否定词缀,合成新词。这样的用词方式更加鲜明地对比出当时人们对于追求自由的向往和社会思想禁锢的矛盾,同时也有力地揭露旧时教会体制下的伪善。

合成词的大量使用一方面增加了《红字》剧情的冲突,提高了文章的可读性,另一方面也从总体上提高了《红字》词汇的平均长度,进而形成了Hawthorne独特的文体风格。

(二)词性分析

词性,作为一种特殊的语法分析单位,通常被认为是衡量某一文学作家或作品文体风格的重要依据 (Milic,1991)。然而,由于不同词性在句子中的功能不同,它们又可以被划分为实词和虚词。英语中的实词 (lexical word 或 content word)主要包括名词、动词、形容词和副词;虚词 (grammatical word 或 function word)主要包括介词、代词、连词和冠词。实词通常具有稳定、完整的词汇意义,而虚词没有完整的词汇意义,但具有语法意义 (胡壮麟,2002)。本研究借鉴胡壮麟的词性分类,对具有词汇意义的实词进行统计分析,结果如图 1 所示:



	名词	动词	形容词	副词
《红字》语料库%	21.31	17.01	8.45	7.24
参照库%	21.87	21.25	4.49	7.12

图 1　两个语料库的实词使用频率统计

通过以上图表可以看出,《红字》语料库中的名词和动词的使用频率低于 BROWNA 文学语料库;而《红字》语料库在形容词和副词的使用频率上则高于 BROWNA 文学语料库,尤其是在形容词的使用频率上,卡方检验的结果表明,两者的差异最为显著 ($\chi^2 = 226.64$, $p < 0.00001$)。这种现象的产生取决于 Hawthorne 的写作特点和小说的内容。首先,Hawthorne 作为"心里小说"的开创者,注重并擅长于剖析人的内心活动 (李宜燮、常耀信,1991)。这意味着 Hawthorne 在文学创作过程中,更加关注对小说人物内心活动细腻而精准地刻画,而为了达到这种目的,Hawthorne 势必会依赖大量的形容性词汇对于人物心理进行描绘。第二,《红字》集中体现了 Hawthorne 的"原罪论"思想,在这部小说中,所有人都是有罪的,每个人都在通过不同的形式来救赎自己的罪恶,而罪恶的救赎的主要途径是心灵上的净化 (常耀信,2008)。因此,在这部小说中,存在着大量关于人性内心罪恶和救赎过程的描写,而这些内心的活动也需要通过大量形容词才能淋漓尽致地表现出来。

(三)主题词分析

主题词分析是语料库研究中常用的文本分析方法。所谓主题词 (key words),指的是与某个参照库中的词汇分布相比,某个特定文本中出现频率显著性高的那些词。因此,提取主题词往往要通过对比某一完整连续文本和一个更大的参照文本 (reference corpus),把语篇中差异显著的词语提取出来,生成一个主题词表。在文学作品分析中,主题词的确立可以帮助人们更加直观地了解文学作品中的基本信息和主要情节。本研究利用 WordSmith Tools 软件中的 KeyWords 功能,以 BROWNA 文学作品语料库作为参照,与《红字》语料库进行比较并提取主题词表如下:

通过表 2 可以看出,《红字》中排在前 20 位的关键词依次为:Hester, Her, Thou, Pearl, Prynne (Hester 的姓), Minister (指 Dimmesdale), Scarlet, Dimmesdale, Child (指 Pearl), Thee (指 Hester), So, She (指 Hester), Thy, Chillingworth, And, With, Letter, Roger (Chillingworth 的姓), Heart 和 Had。这些词汇可以反映出作品的一些基本信息:

首先, 词表中的 Heste, Prynne, Thee 和 She 都是用来表述同一个人物,且这四个词的关键性 (keyness) 都较高;频次紧随其后的是 Pearl, Dimmesdale 和 Chillingworth,这四个人恰好是《红字》中的四个主要人物,可见,故事是围绕着这四个人展开的。其次,通过进一步观察主题词表可以分析出,作品是围绕着红字 (Scarlet Letter) 这一主题进行阐述,并且这个两词在文中出现的频率都很高,这点印证了 Hawthorne"象征主义"的写作手法。此外,heart 一词也出现在主题词表中,说明了小说中充满了大量有关心理活动的描写。最后,通过观察主题词表发现,小说中第二人称的表达方式均选用了古英语形式 (thee, thy, thou),由此可以推断出故事的发生年代并且暗示出小说体裁比较严肃。

表 2 《红字》主题词词表

	Key word	Freq.	%	. Freq.	RC. %	Keyness	P
1	HESTER	360	0.52	3		2,214.03	000000000
2	HER	945	1.37	3,429	0.23	1,650.57	000000000
3	THOU	238	0.34	16		1,368.11	000000000
4	PEARL	221	0.32	12		1,286.43	000000000
5	PRYNNE	150	0.22	0		936.44	000000000
6	MINISTER	155	0.22	70		694.98	000000000
7	SCARLET	105	0.15	1		644.22	000000000
8	DIMMESDALE	102	0.15	0		636.71	000000000
9	CHILD	191	0.28	242	0.02	620.06	000000000
10	THEE	95	0.14	17		499.16	000000000
11	SO	405	0.59	2,194	0.15	477.80	000000000
12	SHE	496	0.72	3,234	0.22	464.76	000000000
13	THY	86	0.12	13		460.99	000000000
14	CHILLINGWORTH	72	0.10	0		449.42	000000000
15	AND	2,286	3.30	30,733	2.05	433.50	000000000
16	WITH	830	1.20	7,729	0.52	430.77	000000000
17	LETTER	124	0.18	158	0.01	401.50	000000000
18	ROGER	77	0.11	16		396.68	000000000
19	HEART	127	0.18	181	0.01	391.67	000000000
20	HAD	658	0.95	5,773	0.39	383.52	000000000

为了能够更好的探索小说《红字》的故事情节,笔者对于主题词表进行进一步处理,生成了主题词表图。主题词表图是根据主题词表计算出各个主题词在文本中的位置分布,进而绘制的主题词分布图。主题词在词图中的出现的先后顺序和密度可以直观地反映出作品主题发展和情节推进情况,进而帮助读者对小说的发展脉络产生直观而具体的认识(张海云、谢群芳,2010)。图 2 展示了《红字》语料库的主题词表图:

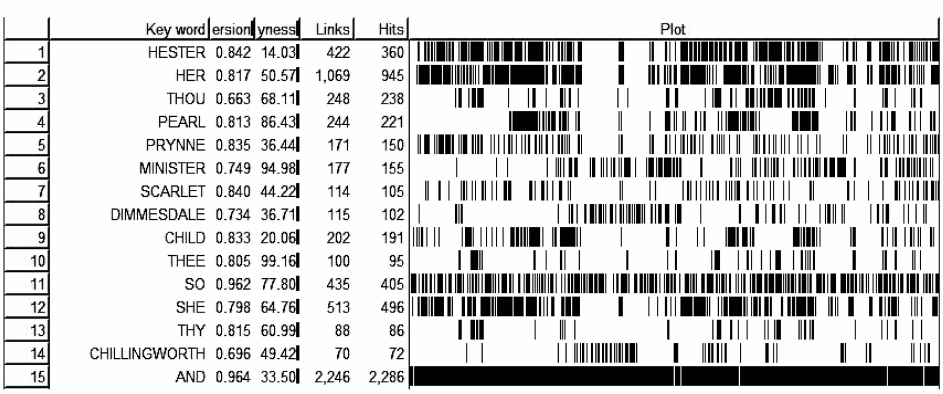


图 2 《红字》主题词表图

通过图 2 可以看出,首先,指代主人公 Hester 的词汇在全文中分布密度较大且比较均匀,这说明了 Hester 这一角色在小说中贯穿始终。细读《红字》,Hester 和丈夫从英国移居到美国波士顿,中途丈夫被印第安人俘虏。Hester 不得不只身到美,到美之后迫于生活,被一青年牧师诱奸怀孕,因此被当地清教徒视为大逆不道,投入监狱,佩戴红字并当众受审,然而,Hester 宁愿受罪,誓死不供出她的情夫,因此被判远离城镇,过着屈辱的生活。在她的孤独生活中,Hester 一边独自抚养她的孩子,一边帮助其他有困难的人,最终,镇中人们终于被她崇高的道德和助人的精神所感动,而曾经令她感到耻辱红字的涵义也变成了天使和道德的化身。

其次,小说另外两个主人公 Dimmesdale (Minister) 和 Chillingworth 最高出现频率都集中在小说的中间部分,并且重合部分较多,这说明两人曾经有着密切的交往和联系。再看小说,Chillingworth 从印第安人手中逃了出来,一直在暗中侦查他的妻子 Hester 的情夫,最终他发现了 Hester 的情夫是当地颇有名望的牧师 Dimmesdale,于是他乔装成为一名医生,接近年轻的牧师,开始了对他疯狂的报复。

再次,小说的第四主人公 Pearl 密集出现的频率主要分为三段,并且这三段都与 Hester 的出现相互重合。回顾《红字》,第六章着重介绍了 Pearl。当 Hester 被判远离城镇的时候,身边只有年幼的女儿 Pearl 与她为伴,然而,由于未受教育以及孤独的缘故,Pearl 行事怪异,而这种怪异深深地刺痛了 Hester 的心;小说第十五章借助 Pearl 一次又一次对于 Hester 胸口前红字追问,深刻地揭示出 Hester 内心之苦;小说结局伴随着 Pearl 的渐渐长大,Hester 身上的红字的涵义也慢慢发生了变化。

最后,通过观察主题词 Scarlet 的分布,可以看出,"红字"贯穿文本始终,充斥了 Hester 的整个生活。这也再一次有力地证明了 Hawthorne"象征主义"的文体风格。

(四)平均句长

平均句长指的是文本中句子的平均长度(由单词数量表示);而句长标准差反映的是文本中句子的长度在平均句长左右浮动的程度,标准差越高说明文本中句子长短变化越大,句式越为灵活,可读性也就越强。平均句长及句长标准差通常被认为是用来衡量作者文体风格的典型因素。表 3 统计了两个语料库的句长信息:

表 3　两个语料库平均句长和句长标准差

	《红字》语料库	参照库
平均句长	21.16	15.02
句长标准差	16.63	16.55

由上表可知,《红字》语料库的平均句长为 21.16,而参照库的平均句长为 15.02,独立样本 t 检验结果表明,这两个语料库在平均句长及其标准差上具有显著性差异($t = 5.71$, $p < 0.01$)。这个结果说明,相对于其他文学作品,Hawthorne 在小说《红字》之中使用了更多复杂的句型,进而使小说在整体上呈现出更为正式、严肃的文体风格。①

这个结果的出现与 Hawthorne 自身的思想矛盾和写作意图有着密切的关系:Hawthorne 作为一名清教徒,一方面深受清教信条的影响,而另一方面,他也对于当时清教徒宗教狂热和不容异端的行为表现出了极大的不满,他认为宗教狂热给当时社会带来了剧大的问题(李宜燮、常耀信,1991)。因此,《红字》的完成更像是 Hawthorne 对于宗教狂热分子的一种宣言,具有强烈的宣传性和批判色彩,他不仅仅将《红字》的阅读人群定位于普通民众,更是将其直接对准了那些受过良好教育的宗教人士,这就导致了 Hawthorne 使用更多的长句,从而提高作品的严肃性和感染力。

五、结语

通过以上对比分析可以看出,基于语料库的文学批评具有独特的优势:它从文本特征、故事情节和艺术特色等角度出发,以客观数据作为支撑,将对文学作品的感性认识和理性研究有机地结合起来;同时,基于语料库的研究方法也为研究作家的语言使用、作品主题的表达、人物形象刻画等方面的特点提供了可靠的量化依据,进而避免了传统文学批评中只注重概念演绎或者生搬某种文学理论进行穿凿附会的弊端(张海云、谢群芳,2010)。

[参考文献]

[1] Chang Yaoxin. A Survey of American Literature[M]. Tianjin: Nankai University Press, 2008.

[2] David. L. H. Some Approaches to Corpus Stylistics [J]. 外国语, 2010,(2): 67-81.

① 有关 CLAWS 词性附码器和 CLAWS7 词性附码集的详细情况请参见 http://ucrel.lancs.ac.uk/claws/trail.html. 和 http://ucrel.lancs.ac.uk/claws7tags.html.

[3] Fowler, R. Style and Structure in Literature: Essays in the New Stylistics[M]. Oxford: Blackwell, 1975.
[4] Hu, Zhuanglin. Linguistics: An Advanced Course Book [M]. Beijing: Beijing University Press, 2002.
[5] Howthorne, N. The Scarlet Letter[M]. London: Harper Press, 2010.
[6] Milic, L, T. Progress in Stylistics: Theory , Statistics, Computers[J]. Computers and Humanities 1991(25): 31-37.
[7] 陈建生,高博. 基于语料库的《诗经》两个英译本的译者风格考察——以"国风"为例[J]. 天津外国语大学学报,2011(4): 36-41.
[8] 蒋原伦, 潘凯雄. 文学批评与文体[M]. 北京:北京师范大学出版社, 2006.
[9] 李晋,郎建国. 语料库语言学视野中的外国文学研究[J]. 外国语, 2010(2): 82-89.
[10] 李宜燮,常耀信. 美国文学选读(下)[M]. 天津:南开大学出版社, 1991.
[11] 马广惠. 基于语料库的小说文体学研究[J]. 常熟理工学院学报, 2005(5): 4-6.
[12] 杨建玫.《警察与赞美诗》的语料库检索[J]. 四川外语学院学报, 2002(5): 56-59.
[13] 杨惠中. 语料库语言学导论[M]. 上海:上海外语教育出版社, 2002.
[14] 张海云, 谢群芳. 基于语料库的文学作品分析——以越南中篇小说《志飘》为例[J]. 解放军外国语学院学报, 2010(3): 57-61.
[15] 张显平. 构建语料库促进英美文学教学改革[J]. 四川外语学院学报, 2007(5): 132-135.
[16] 周艳丽, 张发祥.《德伯家的苔丝》的语料库检索分析[J]. 河南科技大学学报(社会科学版), 2008(4): 62-64.

(责任编辑:杨 睿)

A Corpus-based Approach to Literary Criticism

——A Case Study of The Scarlet Letter

GAO Bo

(*School of Foreign Languages*, *Tianjin University of Science and Technology*, *Tianjin* 300222, *China*)

Abstract: Corpus linguistics, as a new and rising discipline, has been commonly applied to the field of literary criticism. By taking the fictions in the BROWNA corpus as a reference, the comparable corpus of The Scarlet Letter is compiled. With the help of the WordSmith Tools 4.0 of corpus retrieval and analysis software, the information such as mean word length, part of speech distribution, key words and mean sentence length is analyzed quantitatively and qualitatively. The result not only verifies Hawthorne's unique writing style objectively but also reveals the plot of The Scarlet Letter clearly.

Key words: corpus linguistics; literary criticism; The Scarlet Letter; Hawthorne