

# 文学语篇中词汇模式构建机制及其动因研究

汪云

(江苏师范大学 外国语学院, 江苏 徐州 221116)

【关键词】文学语篇; 词汇模式; 语料库

【摘要】文学研究多关注经典作品, 较少涉及非经典作品, 致使研究整体性和科学性存在不足, 而借助语料库方法进行文学分析, 则可以揭示作家整体语言风格。本文通过对马克·吐温作品中词汇模式的研究发现: (1) 较同时代作家, 其用词更加富有变化, 行文中体现出“注重背景及人物描写, 轻对话叙事”的特点; (2) 主题词分析及关键主题词搭配分析揭示其偏爱描写男性儿童, 反映出作者借助“儿童观察世界”的独特视角; 女性角色低频出现, 且多为传统女性, 她们的存在仅为衬托主人公的男子气概且富历险精神; 主人公人名高频搭配词多含消极语义, 反映出作者笔下人物身处社会底层, 受教育程度低下。

【中图分类号】I045

【文献标识码】A

【文章编号】2095-5170(2015)03-0045-06

## 一、引言

文学语篇是语言运用中最主要的形式之一, 如果语言研究不涉及文学语篇, 那么此类语言描写就缺乏系统性<sup>[1]</sup>。美国文学研究最具普遍性的范式, 通常是从具体作家入手, 深入考察其代表作品。但近年来文学研究也开始有意识地借鉴语言研究的理论方法, 强调语言学、翻译或跨文化交流的课题, 强调语言的运用<sup>[2]</sup>。语言学视角下的小说研究, 有助于小说语言特点的挖掘, 并依此揭示作者在语言技巧选择时的目的和用意<sup>[3]</sup>。例如, Lecch & Short (1981) 就曾指出, 作者的心理特点与其作品中语言类的频数信息关系密切, 常见或非常见语言符号的使用与作者经历有密切且直接关系<sup>[4]</sup>。因而, 20 世纪 90 年代以来, 越来越多的学者采用语料库语言学方法进行文学语篇中各语言层面的分析, 使得文学研究趋向于理性和感性结合, 并逐渐走出传统文学作品研究范式的桎梏。

文学文本的语料库加工与研究方法逐渐进入国内学术视野, 但国内对于大规模文学语料库的研究并不多见, 并且当前的“研究内容局限, 偏食

现象严重, 许多非经典作品尚无人问津, 致使研究缺乏整体性和科学性”<sup>[5]</sup>。此外, 传统的语篇分析是个案式的、动态的、从微观到宏观、意义和形式互动的定性分析; 而运用语料库手段进行的语篇分析则是批量式的、静态的、从形式到意义、微观与宏观互动的量化研究。两者的结合将更有利于篇章文体的分析<sup>[6]</sup>。本文运用语料库方法对自建马克·吐温作品集语料库进行量化解读, 分析其词汇构建模式及其动因。通过语料库语言学强大的加工、分析以及统计工具, 对文学文本进行理性的定量分析, 再结合传统的文学文体学分析方法, 从而对文学语篇进行更为详实、科学和理性的描述、分析和阐述。

## 二、国内外研究现状

计算机软件应用于文学分析, 有助于发现文学批评家直觉未注意到的典型文体特征<sup>[7]</sup>。国外对文学语篇的语料库研究, 大多涉及三个领域: (1) 文学语篇的总体语言特征描写, 例如 1981 年 Mick Short 和 Geoffrey Leech 合著了 *Style in Fiction: A linguistic introduction to English fictional prose*<sup>[8]</sup>, 对文学作品中作者所采用的语言表述形式进行了归纳总结; 此后, Short 和

【收稿日期】2014-11-15

【基金项目】本文系江苏省哲学社会科学项目“文学语篇的语料库文体学研究”(项目编号: 11YYD021) 的阶段成果。

【作者简介】汪云, 女, 江苏泗阳人, 江苏师范大学外国语学院讲师。

Culpeper (1997) 选用 40 篇各约 2000 词的文本样例<sup>[9]</sup>, 借助语料库方法, 旨在验证 1981 年 Mick Short 和 Geoffrey Leech 总结的言语与思想表达方式。上述研究虽借助于语料库对不同类型的文体进行了对比, 但语料库规模偏小, 并且当时标注的准确率偏低, 这在一定程度上影响了数据的可靠性。(2) 作品主题的语言表达方式研究, 相关研究包括 Louw (1993)<sup>[10]</sup>, Stubbs (2005)<sup>[11]</sup>, Inaki & Okita (2006)<sup>[12]</sup> 以及 O'Halloran (2007)<sup>[13]</sup>, 但上述研究都只针对单个文本进行, 仅能说明该文学作品某个方面的语言特点, 如果要反映文学语篇词汇语法各层面的语言特点, 则需要从词语搭配、类联接、语义倾向和语义韵几个层面做全面的分析。(3) 作家风格研究, Burrows (1987) 开启了采用多变量统计手段进行文本量化分析之先河, 他使用主成分分析法对奥斯汀作品 30 个常用词进行了分析<sup>[14]</sup>。Hoover (2007) 对 Henry James 语言风格的历时发展变化进行了探究, 发现该作家前期与后期作品中体现出两种完全不同的写作风格, 并认为可能存在五种次风格 (sub-styles)<sup>[15]</sup>。Tabata (2002) 则采用词类分布对应分析法 (correspondence analysis) 探讨了狄更斯语言风格的变化<sup>[16]</sup>。以上研究多集中于高频词频数统计或对比, 缺乏对文学文本中主题词及关键主题词的搭配词分析, 而以上视角可揭示文学语篇中词汇模式的全貌, 进而, 结合文学批评理论, 将极大地丰富作家的语言风格研究。

目前, 国内文学语篇的语料库研究尚处于引进阶段<sup>[17][18][19][20][21]</sup>, 少数学者运用语料库方法对单个文本进行了研究<sup>[22]</sup>, 而文学实证研究还需进行更深入、更系统的研究。本研究在借鉴前人研究的基础上, 对文学语料进行标注, 分析并解读单个作家众多作品中所反映出的整体词汇模式构建机制及其动因。

### 三、语料与研究方法

#### (一) 语料来源

特定语篇或作家作品与同时代其他作家作品的共时多维度对比分析可揭示作家语言风格特点。本文以马克·吐温作品为研究范例, 借助语料库工具对其作品词汇模式进行解读。首先, 收集 19 世纪中叶到 20 世纪中叶与马克·吐温同时代的英美作家小说 120 余部, 存为纯文本格式, 使用 PowerGREP 借助正则表达式对文本进行清理, 从而建成 500 余万词英语名著参照语料库; 此

外, 收集马克·吐温经典及非经典中篇及短篇小说 20 余部, 建成观察语料库, 共计 120 余万词。为实现语料增值及提高词汇句法提取的准确性, 使用 CLAWS 7 对两个语料库进行了词性赋码。

#### (二) 研究问题

本文首先对观察语料库中的词汇总体特征、词汇模式的构建进行数据统计, 然后借鉴文学批评理论, 讨论分析上述词汇模式构建的动因。基于上述研究目的, 本文拟回答以下问题:

1) 与同时代作家作品相比, 马克·吐温作品呈现何种词汇总体特征?

2) 与同时代作家作品相比, 马克·吐温作品中词汇模式构建特点及其动因是什么?

#### (三) 研究方法

针对第一个问题, 本文运用 WordSmith 对参照语料库与观察语料库的标准类形符比、平均词长、平均句长及词汇密度进行比较, 以分析马克·吐温作品的词汇总体特征。针对第二个问题, 提取词汇层面的相关数据, 首先借助 WordSmith 中 KEYWORDS 提取主题词, 然后运用 WORDLIST 中 RELATIONSHIP 功能对关键主题词强搭配词及其语义进行分析, 描述并阐释马克·吐温作品中词汇模式的特点及其原因。

### 四、分析与讨论

#### (一) 文本词汇总体特征

数据描述及理论阐释相结合是语料库文体学提倡的研究方法, 两者形成非常好的互补。本研究中, 数据描述在前, 剖析阐释在后, 进而揭示文学语篇中的词汇总体特征。通常来说, 语料库中所反映的总体特征主要体现在标准类形符比、平均词长、平均句长及词汇密度等几个方面。

表 1 文本总体特征

	观察语料库	参照语料库
形符数	1,190,401	5,178,056
标准类形符比	44.75	41.74
平均词长	4.19	4.21
平均句长	20.07	15.38

表 1 显示, 观察语料库中的标准类形符比参照语料库高, 说明马克·吐温词语使用丰富度高于同时代其他英美作家平均水平, 相对多样且富有变化。平均词长方面, 观察语料库与参照语料库水平非常接近, 说明同时代作家在选词方面类似。进一步挖掘数据, 可以发现两个语料库中大都为 6 个字母以内的词, 所占频率分别为 84.8%

和 84.32%，而超过 15 个字母的词在两个语料库中所占比例非常小，平均每一万个词的频数分别为 1.47 和 1.53，总体看来参照语料库与观察语料库在词汇层面难易程度相近。

平均句长方面，马克·吐温作品中句子长度远高于参照语料库，说明马克·吐温使用了较为复杂的句式。根据当代英语平衡语料库 Crown 和 CLOB 中四种语体的平均句长：23.60(学术)，21.04(通用)，20.55(新闻)，13.08(小说)，可以看出马克·吐温作品中平均句长高于英语平衡语料库中小说的平均句长，但低于其他三类语体，而参照语料库平均句长与平衡语料库中小说平均句长近似。结合标准类形符比和平均句长两方面数据，可推断马克·吐温作品在部分语言层面的难度水平高于小说语体平均水平。

通常说来，小说中对话比重越大，平均句长越

短；背景信息描述越多，平均句长相对较长。统计发现，在参照语料库及观察语料库中，含对话句子频数在句子总数中所占比例各占 12.9% 和 10.8%，对比分析说明马克·吐温作品中对话所占比重低于同时代英美作家作品中的比重，可推测马克·吐温更加注重背景及人物性格描述，而轻对话叙事。

词汇密度是文本信息量大小的衡量标准之一，通常体现在实义词在文本中所占的比例，而实义词包括名词、实义动词、形容词和副词四类<sup>[23]</sup>。此外，词汇密度也是区别语体正式程度的一个标准，语体越正式，词汇密度越高；反之，语体越不正式，越接近自然产出语言，词汇密度越低。例如，Ure 研究发现口语中词汇密度在 40% 以下，书面语的词汇密度则在 40% 以上<sup>[24]</sup>。

表 2 参照语料库与观察语料库中实义词频数及比例

	参照语料库		观察语料库		对数似然率	p 值
	频率	百分比	频率	百分比	(LL)	
名词	1131760	21.86	262282	22.03	13.71 * * *	0.00000
动词	650082	12.55	153036	12.86	69.32 * * *	0.00000
形容词	287256	5.55	74558	6.26	850.89 * * *	0.00000
副词	291314	5.63	75788	6.37	897.71 * * *	0.00000
词汇密度		45.33		47.52		

此处星号为对数似然率显著性标记。\* 代表在 p 值小于 0.05，\* \* 代表

在 p 值小于 0.01，\* \* \* 代表在 p 值小于 0.001。

表 2 所示，观察语料库中各类实义词所占比例都略高于参照语料库。实义词在观察语料库中所占比例为 47.52%，而参照语料库的比例为 45.33%，总体差距不大。通过比较参照语料库与观察语料库各词类差异时发现，对数似然率都高于显著差异临界值( $p < 0.001$ )，因而在实义词的使用上，两个语料库间具有统计学意义上的显著差异。结合表一标准类形符比可推测，在一定程度上马克·吐温用词的丰富度高于同期作家，传达信息量更大，这也在一定程度上增加了读者的阅读难度。

## (二) 马克·吐温作品词汇模式构建及其成因

### 1. 主题词分析

本文在提取主题词时，运用了 WordSmith 中的 KEYWORD 功能。此工具对分析作家作品中的词汇模式具有优势，因为“主题词研究是研究文本内容和文本语言特征差异的重要手段”<sup>[25]</sup>。通过观察主题词词表(正主题词和负主题词)，可以得到观察语料库中词汇特征的相关数据，关键值越高说明该词与作品内容关联性越高，主题词关

键值为负值时说明其与作品内容相关度极低。

表 3 观察语料库主题词表

正主题词	Keyness	负主题词	Keyness
EN	1,479.46	HER	-6,565.36
JIM	1,432.33	SHE	-4,453.59
TOM	1,270.81	MRS	-2,203.30
WARNT	1,238.11	LOVE	-671.49
HUCK	1,231.65	LADY	-667.25
RIVER	927.99	COLONEL	-570.7
COULDN'T	732.26	DEAR	-435.71
DIDNT	626.26	FATHER	-391.79
DUKE	475.23	MOTHER	-356.71
MISSISSIPPI	471.53	WOMAN	-355.11
RAFT	380.8	HELL	-305.53
DOLLARS	373.25	HUSBAND	-294.73
MIGHTY	356.63	MARRIED	-277.17
NIGGER	355.77	LOOKED	-260.91
DEY	351.77	GIRL	-260.82

表 3 中正主题词反映出马克·吐温作品中的情节大都围绕 Jim, Tom 及 Huck 展开，地点则在密西西比河流域，这与马克·吐温自身生活经历有关，他曾在蒸汽机船上担当导航员，接触过形形色

色色的人物,对周围事物熟悉,同时密西西比河是当时美国资本主义发展的缩影,因而小说场景大都围绕这条河流展开。

主题词 WARN'T, COULDN' 以及 DIDN'T 为否定助动词,均为过去式形式。结合 3 个词的索引行,可发现其多出自 3 部作品,每千词频数及分散率远远高于其他作品(见表 4)。同时,这些词多出自 Tom, Jim 和 Huck 之口,尤其是 WARN'T 的使用,此词在正式语体属语法错误,而在作品中多次出现在主人公话语中,这充分说明其笔下主人公身处社会底层,缺乏正规教育。同时,否定词在主人公话语中多次复现也揭示出三位主人公生活不如意、不幸福,反映出他们对社会的不满及对等级制度的厌恶。另两个主题词的出现(DUKE 公爵和 DEY 总督)则反映出资本主义国家人与人之间距离的疏远以及阶级差异鸿沟的存在。结合马克·吐温作品的创作年代,可以看出,当时正处在资本主义发展时期,物质财富得到极大提升,而公民教育却很落后,底层阶级受到多种不平等对待,因此作家试图借人物角色之口对上述现象进行讽刺。

表 3 中负主题词多为与女性相关的代词、名词或是称谓,如 HER, SHE, Mrs. 等等,这充分说明马克·吐温笔下对女性角色涉及少于同时代其他作家。结合正主题词可以看出其偏爱男性儿童主人公,所以马克·吐温作品中极少以女性为主角。Fishkin (1995)论述马克·吐温笔下女性角色时引用了部分学者的观点,“William Dean

Howells 在 1910 年认为‘马克·吐温对女性缺乏了解,因此其笔下女性角色远不及儿童角色及男性角色那么鲜活’;Mary Ellen Goad 在 1971 年评论说:‘马克·吐温缺乏成功塑造女性角色的能力,其笔下女性往往比较呆板和不切实际’;Judith Fryer 在 1976 年认为‘马克·吐温根本没有创造过任何典型女性形象’,总体说来,马克·吐温在作品中不乐于描述女性角色”<sup>[26]</sup>。Harris (2002)论述女性角色在作品中的地位时指出“马克·吐温笔下女性多为传统类型,她们的存在仅仅是为了衬托男性主人公的男子气概,显示其笔下男性主人公善于打破常规及习俗的能力”<sup>[27]</sup>。史永红、张丽娟(2012)则认为,“马克·吐温作品暴露了作者生态女性意识的矛盾性”<sup>[28]</sup>。

此外, LOVE 一词在负主题词词表中高居第四位,这说明该词在马克·吐温作品中非常鲜见(其作品中仅出现 228 次),并反映出小说中人物缺乏关爱以及资本主义社会人与人之间的冷漠。FATHER 和 MATHER 两词在负主题词中出现,说明主人公缺乏父母关爱,正如 Harris (2002)所说“马克·吐温作品中父亲角色要么过世,要么无为,所以其作品中主人公多为儿童”<sup>[29]</sup>;HUSBAND 和 MARRIED 二词在负主题词中的出现则体现了马克·吐温笔下人物多为孩童,极少涉及婚姻家庭。结合马克·吐温一生当中的不幸,可以发现作家借助儿童视角观察美国‘文明’社会,从外界审视这个自认为文明进步、平等民主的社会<sup>[30]</sup>。

表 4 观察语料库否定词频数及其分散率

	哈克贝里·芬历险记		汤姆·索亚在国外		侦探汤姆·索亚	
	每千词	分散率	每千词	分散率	每千词	分散率
WARN'T	2.66	0.901	1.92	0.846	1.98	0.696
DIDN'T	3.15	0.909	2.57	0.875	3.09	0.869
COULDN'T	1.97	0.919	3.16	0.848	2.23	0.806

## 2. 主题词搭配词检索分析

表 3 显示,马克·吐温作品中主人公人名排在主题词表的前几位。为进一步认识各人物形象,运用 WordSmith 的 WORDLIST 中 RELATIONSHIP 功能检索与主人公人名共现语境中

的名词、形容词、动词和副词,将共现频数高且搭配强度大的词和短语进行分类、归纳和分析,能够帮助读者获取作者用于描述人物外貌、活动、性格、心理活动等方面的有力数据,为以往文学作品的感性研究提供理据。

表 5 观察语料库中关键主题词强搭配词表

Tom 搭配词	对数似然率	Jim 搭配词	对数似然率	Huck 搭配词	对数似然率
SAWYER	1679.80	NIGGER	108.67	FINN	690.28
SAID	1182.40	MARS	96.30	TOM	213.35
MARS	1179.01	LEM	83.33	DON'T	205.90
CANTY	519.72	WATSON	77.36	AIN'T	92.45
HUCK	213.35	HUCK	47.57	EN	85.53

BECKY	164.26	WARN'T	44.18	JIM	63.94
JIM	159.91	RAFT	40.45	CAN'T	60.34
MARSE	181.51	BEN	38.26	LOOKY	54.28
AIN'T	139.97	FREE	34.71	ILL	44.57
LOOKY	78.50	AWFUL	28.36	MIGHTY	31.20

表5显示,马克·吐温笔下三位主人公人名强搭配词多具消极语义倾向,如 ain't, nigger, awful, warn't, don't, can't, ill 等词。Tom 和 Jim 搭配词中 mars 以及 marse 是美国南部黑人指代主人用语,通过对包含 mars 和 marse 的索引行进行解读,发现大部分都是 Jim 对 Tom 的称呼,结合 Jim 的搭配词 nigger(黑鬼,对黑人的蔑称),可以看出马克·吐温在语言方面大量融合、提炼、吸收方言土语,运用密苏里黑人土语、西南边疆地区方言、密西西比河流域普通方言<sup>[31]</sup>。

此外,awful 一词与主人公的高频共现,反映了作家创作时期黑人和白人所处地位不同,阶级差异的普遍存在以及处于社会底层人们生活的艰辛与痛苦。AIN'T 作为 am not, are not, is not, has not 和 have not 缩略语用法已有很长历史,但使用 AIN'T 逐渐被认为是文盲标志,这也说明马克·吐温笔下的三位主人公地位低下,缺乏良好的教育。结合三人话语的索引行,可以发现诸多不符合语法规则的词语或语句,马克·吐温正是运用这些词语间接向读者描述主人公缺乏教育,文明程度相对低下,这与资本主义物质财富的快速增长形成鲜明反差,借此对社会进行批判。

Jim 搭配词中 raft 和 free 两词的出现,可以反映出 Jim 在试图摆脱白人控制,获取自由而乘木筏溯河而进时所作出的努力,体现了主人公不满现状,厌恶庸俗保守的生活,厌恶窒息人的社会制度,追求解放,追求自由、开拓新天地的精神。三位主人公的精神境界,正是 19 世纪美国普通人民西部拓荒时期精神境界的缩影。

## 五、结语

本文借助语料库分析工具,对马克·吐温与同时代作家的作品中的词汇模式进行了对比,并分析了马克·吐温作品中整体词汇模式的构建特点及其成因。研究发现:(1)在标准类形符比和词汇密度两类数据上,观察语料库高于参照语料库,反映出马克·吐温的作品在词汇层面的复杂度高于同时代作家作品的平均水平;(2)通过主题词分析发现,马克·吐温作品中人物主要涉及 Tom, Jim 和 Huck,而上述人物普遍处于社会底层,受到正规教育较少,缺乏关爱;此外,其作品中对于女性角色涉及很少,尤其是小说中主人公极少为女性,并且女性角色的出现大都为了衬托男性主人公的男子气概;通过对主题词的强搭配词进行搭配分析发现,主人公人名的搭配词普遍具有消极语义倾向,反映了作家创作的时代特点,即处于社会底层的人对当时社会的反感及厌恶,表现出强烈摆脱束缚获取自由的倾向。

基于语料库的文学作品分析方法“虽然在语言特征判断方面无重大突破,但实施起来却快捷、准确,省时省力,而且证据充实,令人信服”<sup>[32]</sup>。建立大规模标注文学语料库,将会使文学语篇的研究更为深入,不仅可体现作家整体语言风格,也能揭示文本的深刻内涵。文学语篇的语言学研究将会对外国文学研究提供新的研究视角和新的理论框架,对于语料库语言学则是新的开拓与尝试,同时为语言学与文学研究纽带的搭建及文学教学新方法的构建奠定了一定的基础。

## 参考文献

- [1] Sinclair, J. *Trust the Text*. New York: Routledge Taylor & Francis Group. 2004, p. 51.
- [2] 杨金才:《从研究范式看美国文学研究的界面性》,《外国语文》,2014 年第 1 期。
- [3] Biber, D. & Conrad, S. Register, *Genre and Style*. Cambridge: Cambridge University Press. 2009, p. 132.
- [4] [8] Leech, G. & Short, M. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. New York: Longman, 1981.
- [5] 杨金才、于雷:《中国百年来马克·吐温研究的考察与评析》,《南京社会科学》,2011 年第 8 期。
- [6] 杨惠中:《语料库语言学导论》,上海外语教育出版社,2002 年版,第 42 页。
- [7] [11] Stubbs, M. "Conrad in the computer: examples of quantitative stylistic methods". *Language and Literature*, 2005, (1).
- [9] Short, M. & Culpeper, J. "Using a computer corpus to test a model of speech and thought presentation". *Poetics*, 1997, 25.
- [10] Louw, B. "Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies". Baker, M. G. Francis & E. Tognini - Bonelli. *Text and Technology: in Honour of John Sinclair*. John Benjamins, 1993, 157-176.

- [12]Inak, A. &Okita, T. "A small corpus—based approach to Alice's roles". *Literary and Linguistic Computing*, 2006, (3).
- [13]O'Halloran, K. A. "Corpus—assisted literary evaluation". *Corpora*, 2007, (1).
- [14]Burrows, J. *Computation into Criticism: a study of Jane Austen's novels and an experiment in method*. Clarendon Press, 1987.
- [15]Hoover, D. "Corpus Stylistics, Stylometry, and the Styles of Henry James". *Style*, 2007, 41(2).
- [16]Tabata, T. "Investigating stylistic variation in Dickens through correspondence analysis of word—class distribution". T. Saito, J. Nakamura & S. Yamazaki. *English Corpus Linguistics in Japan*. Rodopi, 2002, 165—182.
- [17]李晋、郎建国:《语料库语言学视野中的外国文学研究》,《外国语》,2010年第2期。
- [18]李涛、王菊丽:《语料库文体学:计算机辅助文学语篇的文体分析》,《外语电化教学》,2009年第1期。
- [19]吕长炫:《语篇的语料库研究范式评介》,《外国语》,2010年第2期。
- [20]卢卫中、夏云:《语料库文体学:文学文体学研究的新途径》,《外国语》,2010年第1期。
- [21]刘靖、黄立波:《〈语料库文体学〉述介》,《外语教学与研究》,2010年第3期。
- [22]马广惠:《基于语料库的小说文体学研究》,《常熟理工学院学报》,2005年第5期。
- [23]Biber, D. et al. *Longman Grammar of Spoken and Written English*. Foreign Language Teaching and Research Press, 2000, 62.
- [24]Ure, J. "Lexical density and register differentiation". G. Perren & J. Trim. *Applications of Linguistics: Selected Papers of the Second International Congress of Applied Linguistics*. Cambridge University Press, 1971, 443—452.
- [25]王立非、梁茂成:《WordSmith 方法在外语教学研究中的应用》,《外语电化教学》,2007年第6期。
- [26]Fishkin, S. "Mark Twain and Women". G. Forrest. *The Cambridge Companion to Mark Twain*. Cambridge University Press. 1995, 52—73.
- [27][29]Harris, S. "Mark Twain and Gender". S. Fishkin. *A Historical Guide to Mark Twain Historical Guides to American Authors*. Oxford University Press, 2002, 163—194.
- [28]史永红、张丽娟:《论生态女性意识及其矛盾性》,《宁夏大学学报》,2012年第3期。
- [30]Robert, F. "Mark Twain: A Brief Biography". S. Fishkin. *A Historical Guide to Mark Twain Historical Guides to American Authors*. Oxford University Press, 2002, 13—51.
- [31]Fishkin, S. "Mark Twain and Race". S. Fishkin. S. Fishkin. *A Historical Guide to Mark Twain Historical Guides to American Authors*. Oxford University Press, 2002, 127—162.
- [32]何安平:《〈语料库研究语言〉导读》,外语教学与研究出版社,2001年版,第28页。

## On the Construction Mechanism of Lexical Patterns in Literary Discourse

WANG Yun

(School of Foreign Studies, Jiangsu Normal University, Xuzhou 221116, China)

**Key Words:** literary discourse; lexical patterns; corpus

**Abstract:** Literary studies mainly focus on classics and master pieces, neglecting non-classics. The unification of corpus and stylistic studies can help to reveal the characteristics of specific authors' language. The analysis of lexical patterns in Mark Twain's works shows that: (1) STTR and lexical density in his works are higher than those in reference corpus, which implies that Twain put emphasis on background description, instead of on dialogues; (2) keyword analysis reveals that heroes are mostly male characters instead of females, who function as foils for young protagonist's adventure; Moreover, words co-occurring with key keywords often carry negative semantic prosody, indicating protagonist's lower class and uneducated state.

[责任编辑:林晓雯]