

基于语料库的《瑞普·凡·温克尔》 多元主题解析

伍勇进

湖南师范大学外国语学院, 湖南 长沙

收稿日期: 2023年1月23日; 录用日期: 2023年2月16日; 发布日期: 2023年2月24日

摘要

华盛顿·欧文在美国文学史上占有重要地位, 被誉为美国文学之父, 但很少有学者对其作品进行定量研究。因此, 本文结合AntConc, Vosviewer和ROST CM6软件, 对其最为知名的短篇故事《瑞普·凡·温克尔》的主题进行基于语料库的数字人文研究。通过语料库的高频词、主题词与社会语义网络从情节和意象方面更加直观地解读该故事历史书写、自然神、反清教以及印第安失落的主题思想。这说明结合语料库方法分析文本能够辅助发现潜在线索, 为文学作品的研究提供了新的可能。

关键词

语料库, 数字人文, 定量分析, 《瑞普·凡·温克尔》, 远读

A Corpus-Based Analysis on Multi-Theme of *Rip Van Winkle*

Yongjin Wu

Foreign Studies College, Hunan Normal University, Changsha Hunan

Received: Jan. 23rd, 2023; accepted: Feb. 16th, 2023; published: Feb. 24th, 2023

Abstract

Washington Irving occupies an important position in the history of American literature and is regarded as the father of American literature. However, few scholars have made quantitative studies on his works. Therefore, this paper combines AntConc, Vosviewer and ROST CM6 software to conduct a corpus-based digital humanities study on the theme of his most famous short story *Rip Van Winkle*. Through the high-frequency words, keywords and social semantic network of the corpus, the theme of historical writing, deism, anti-Puritanism and Indian loss are interpreted more intuitively.

tively from the aspects of plot and image. This research shows that corpus-based analysis can help to find potential clues and provide a new possibility for the study of literary works.

Keywords

Corpus, Digital Humanities, Quantitative Analysis, *Rip Van Winkle*, Distant Reading

Copyright © 2023 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

长期以来,内省式研究方法在文学研究中独步天下,只是到了20世纪80年代初,语料库方法才开始应用于文学研究之中。内省式方法是指研究者依据自己的直觉和判断,就具体文学事实或文学现象提出理论假设,并选择少量典型的案例加以论证的方法。内省式方法的应用往往使得具体研究结论具有一定程度的主观性和片面性[1]。随着数字信息媒介的发展,人类社会逐步进入大数据时代,数字媒介和人文学科的结合与交叉,重构了传统人文研究的知识谱系与认知概念,从而引发了研究范式的革新,将人文学科引向了更为系统、更为多元、更为复杂的新研究领域,自此数字人文应运而生。董洪川指出数字人文视阈下的外国文学研究范式应以定量辅助定性研究,将社会科学的定量研究方法与人文学科的定性研究方法结合起来,对文学文本的既有信息进行检索、提炼和呈现,并从多层面、多维度生成二次信息、提出新的问题,得出新的结论[2]。鉴于此,本文选取华盛顿·欧文的代表作品《瑞普·凡·温克尔》为研究对象,结合数字人文的“远读”技术对该短篇故事进行文本挖掘和定量分析,最后辅以内省式的文本细读试图发掘故事的潜在主题。

2. 语料来源与数据处理

本文选取 *Rip Van Winkle* 原文及2005年万紫翻译的《欧文短篇小说选》中对应篇目的中文译文为数据来源,采用语料库检索软件 AntConc3.5.8,可视化软件 Vosviewer 及文本内容挖掘软件 ROST CM6 用于文本高频词和主题词表生成,前人文献主题的可视化以及该故事的社会语义网络图谱绘制。为避免由文本中大量存在的诸如“go, went, mountain, mountains”等因屈折变换导致的同义表达及在各类文本中都高频出现的如“of, a, the, and”等虚词或助动词所带来的统计偏差,本文使用 Lemma List (词元词表)和 Stopword List (停用词表)进行归一化和去停用词的文本预处理。考虑到欧文的出生时间(1783年4月3日)和 *Rip Van Winkle* 的面世时间(1819年),本文共搜集了1778~1817年间创作的14部英文小说(细节目录见表1),应用正则表达式[A-Za-z0-9-]+ 统计文本字数共2,203,222词(不含标点符号),作为参照语料库(reference Corpus),*Rip Van Winkle* 原文作为观察语料库(observed corpus)。该参照语料库为观察语料库的300余倍,前者较后者足够大,主题词统计结果可以较为精确地反映出语篇层面词语的分布与文本主题的关系以及词语之间的关系。

3. 高频词分析

本文将《瑞普·凡·温克尔》原文处理成 utf-8 格式的 txt 文本后导入 AntConc3.5.8 中,在 AntConc 中分别选择 Tool Preferences-Word List-Lemma List 和 Tool Preferences-Word List-Word List Range-Use a stoplist below-Add Words From File 导入 Lemma List 和 Stop List,最后应用 AntConc 的 Word List 功能生成原文词频表,如下表2。

Table 1. Details of self-built corpus**表 1.** 自建语料库详情

时间	作品名	时间	作品名
1778	<i>Evelina</i>	1796	<i>Camilla</i>
1782	<i>Cecilia</i>	1797	<i>The Italian</i>
1789	<i>The Castles of Athlin and Dunba</i>	1811	<i>Sense and Sensibility</i>
1790	<i>A Sicilian Romance</i>	1813	<i>Pride and Prejudice</i>
1791	<i>Romance of the Forest</i>	1815	<i>Emma</i>
1794	<i>Lady Susan, the Watsons, Sandit</i>	1817	<i>Northanger Abbey</i>
1794	<i>The Mysteries of Udolpho</i>	1817	<i>Persuasion</i>

Table 2. High-frequency wordlist generated by AntConc**表 2.** AntConc 生成的高频词表

Rank	Freq	Lemma	Lemma Word Forms	Rank	Freq	Lemma	Lemma Word Forms
1	64	Rip	Rip 64	17	11	eye	eye 3 eyes 8
2	35	Van	Van 35	18	11	poor	poor 11
3	31	old	old 31	19	11	strange	strange 11
4	31	Winkle	Winkle 30 Winkles 1	20	10	change	change 3 changed 5 changes 2
5	28	mountain	mountain 15 mountains 13	21	10	day	day 9 days 1
6	25	man	man 22 men 3	22	10	gun	gun 10
7	21	village	village 20 villages 1	23	10	hat	hat 9 hats 1
8	18	time	time 13 times 5	24	10	house	house 7 houses 3
9	17	little	little 17	25	10	wife	wife 7 wives 3
10	17	long	long 16 longer 1	26	10	year	year 1 years 9
11	16	great	great 16	27	9	child	child 4 children 5
12	16	name	name 15 names 1	28	9	country	country 9
13	13	dog	dog 11 dogs 2	29	9	Dame	Dame 9
14	13	head	head 13	30	9	door	door 6 doors 3
15	12	rock	rock 6 rocks 6	31	9	Dutch	Dutch 9
16	12	tree	tree 8 trees 4	32	9	story	stories 3 story 6

3.1. “Old” ——怀旧与历史编织

通过高频词表可以发现“old”这一形容词在文中出现的频次竟同主人公 Rip Van Winkle 出现的频次大致等同，可以说“old”隐喻了作者在文本创作过程中对“过往”、“历史”以及“更迭”有意或无意的关注。利用 AntConc3.5.8 的情节检索(Concordance Plot)功能，可以发现“old”实际上贯穿了故事始终(见图 1)，尤其在故事的后半段，也就是瑞普苏醒回村部分。文中反复出现修饰词“old”，同时作者在故事结尾刻意将瑞普的晚年生活安排成“恢复了他往日的行径和习惯；不久他又找到了许多老朋友，不过这一班人都已经到了风烛残年，因此他宁愿跟晚一辈的人交朋友，不久他就得到他们的爱戴。他在家里也没有什么事情可做，而且已经到了逍遥闲散的幸福年龄，因此，他又终日坐在旅店门口，大家都尊他为

村中的老前辈，把他看作一部活的“战前”旧时代的历史” ([3], p. 726)，可见瑞普醒后的生活，或者说是美国独立后的生活较之前并无二致，“old”的高频出现形成了怀旧的复调，一方面，流露出作者对过往田园隐逸生活的怀恋，而另一方面，也借瑞普之口，表达了作者对历史更迭与政治认同的逃遁与保守。



Figure 1. Concordance Plot of “old”

图 1. “old” 情节索引

结合 Concordance 检索(见图 2)，发现“old”在原文中大多用在荷兰先祖的幽灵之前。换句话说，文中高频词“old”形成的怀旧复调实际上是与荷兰先祖的鬼魅形象反复勾连的，“old”指涉的历史性与荷兰早期移民的历史记忆存在共线性，文中对于美利坚历史的书写是经由荷兰移民幽灵的记忆编织起来的。朱迪丝·理查森指出“幽灵作为一种文化记忆形式具有反常的吸引力，它能够从阴影中编织出历史感，并表达和消除一个建立在殖民主义掠夺基础上的主要由陌生人组成的国家所固有的社会焦虑” [4]。美利坚生来就是民主共和国的“成年人样态”，而由于缺乏“孩童期”的成长经历，美国人身份就成为一种基于鬼神传说、历史叙事与逸闻趣谈的杂糅建构。在此建构过程中，作为先驱者的欧文采用幽灵故事赋予荷兰人以美国本土或至少是归化了的地方精神的特权，使他们成为历史记忆的核心，于是瑞普祖先的地方记忆就通过幽灵这个媒介关联起了美利坚民族身份的建构[4]。这或许证明，欧文相信纽约历史的根基孕育在荷兰传统之中，这种传统指引着纽约人乃至美国人为自己的身份寻根，从而构成了书写美利坚国家历史不可或缺的源泉。

Hit	KWIC
1	, not one of which he recognized for an old acquaintance, barked at him as he passed. The
2	on his favorite topics; whereas he found the old burghers, and still more, their wives, rich in
3	ikeness, promised to inherit the habits, with old clothes, of his father. He was generally seen
4	his father had once seen them in their old Dutch dresses playing at ninepins in the hollow
5	point on which he always remained flighty. The old Dutch inhabitants, however, almost universally ga
6	belief, for I know the vicinity of our old Dutch settlements to have been very subject to
7	the country had thrown off the yoke of old England—and that, instead of being a subject
8	, also, about securing the gun, and keeping the old fellow from doing mischief; at the very suggestio
9	's appearance. He was a short, square-built old fellow, with thick bushy hair, and a grizzled
10	clean well-oiled fowling-piece, he found an old firelock lying by him, the barrel encrusted with
11	group reminded Rip of the figures in an old Flemish painting, in the parlor of Dominie Van
12	ions, as a book of unquestionable authority. The old gentleman died shortly after the publication of h
13	hend the question; when a knowing, self-important old gentleman, in a sharp cocked hat, made his
14	the papers of the late Diedrich Knickerbocker, an old gentleman of New York, who was very curious
15	to be the commander. He was a stout old gentleman, with a weather-beaten countenance; he
16	windows, some of them broken, and mended with old hats and petticoats, and over the door was
17	last I saw him, was a very venerable old man, and so perfectly rational and consistent on

Figure 2. KWIC of “old” in AntConc

图 2. “old” KWIC 检索

3.2. “Mountain” ——自然的神性与恒常

通过高频词表可知“mountain”是全文出现频次最高的名词，是故事情节发展的重要组成部分和典型意象。通过比对“mountain”和主人公“Rip”的出现情况(分别见图 3，图 4)可以发现两者之间存在绝

对的重叠和共现，“mountain”参与了主人公瑞普从“入山遇仙”到“饮酒入眠”再到“醒后返村”最后发现“沧海桑田”的整个过程，同时值得注意的是，“mountain”在文本中出现的范围居然比作为故事主人公“Rip”出现的范围更广。比对图3和图4，可以发现“mountain”就好像一个闭环将主人公“Rip”的整个情节囊括其中，而“mountain”独有两段则是故事开篇和尾注部分。换句话说，“mountain”在“Rip”的故事发生前就已出现，而在“Rip”的故事落幕后“mountain”也依然存在于文本的边缘部分——尾注中，从这一角度来看，“mountain”就像一个恒常的存在，不以人的意志为转移。对于这一现象的解释或许可以从以下两个角度进行推论和解读。



Figure 3. Concordance plot of “mountain”

图 3. “mountain”情节索引

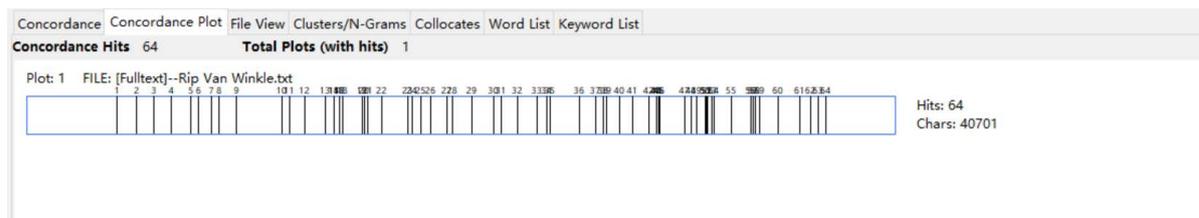


Figure 4. Concordance plot of “Rip”

图 4. “Rip”情节索引

作者在故事开篇就重点描述了神奇的丛山(fairy mountains)，提及群山“威凌四周的乡村”(lording it over the surrounding country)，暗示群山是具有神性的，凌驾于凡尘之上，按地位尊卑，也当先谈及山林，而后再描写凡尘的人与事；故事开篇提到“四季的每一转换，气候的每一变化，乃至一天中每一小时，都能使这些山峦的奇幻的色彩和形态变换，远近的好主妇会把它看作精确的晴雨表”(714)，透露出具有神性的群山是发动“变换”的祖地。在原文中，河流发源于群山，这也使得哈德逊河流(Hudson)也具有着神性。当瑞普“不知不觉地爬上了卡兹吉尔丛中一个极高的峰顶”往下眺望时，发现“here and there sleeping on its glassy bosom and at last losing itself in the blue highlands”。这段描述直接暗示了这片山林具有创造镜像(复制)和令人沉睡的神奇魔力。瑞普回村后，“看见一个和他自己上山时长得一模一样的人；神气也是那么懒散，身上当然也是同样的褴褛。可怜的瑞普现在完全搞糊涂了。他甚至对自己也有点怀疑了：不知道他究竟是瑞普呢，还是另外变了一个人”([3], p. 723)。山林发动“镜像和复制”的魔力，在瑞普进入异时空陷入沉睡后，又在现实世界选中瑞普的儿子复刻出新的“瑞普”，完成“瑞普”基因的接续和传承。瑞普进入洞穴中的圆形剧场可以看成是重回母体的子宫。如果说瑞普喝下了“死人之酒”意味着死亡(非事实性)，那么他在山谷里的睡眠就是重归母体的暗示，而他的醒来则意味着重生。重生前的他因为喝了“死人之酒”而融入了荷兰祖先们存在的空间，故而不受凡常时间的控制；而重生后，瑞普需要回到他原来生活的环境，就必须重新受凡间时间的控制，也就是重回群山这一“变换”的祖地。因此他的一夜昏睡等于凡间的二十年。当瑞普下山回村时“他开始怀疑他和这周围的世界，是不是都中了魔法，这明明是他的家乡，他离开了不过一天”([3], p. 721)。而让瑞普打消这一疑虑，确信自己并未走错的理由却是因为他看见“那里是巍然耸立的卡兹吉尔丛山——远远流着的是银色的哈得逊河——每

一重小山，每一个溪谷，都和往日完全一样——瑞普心里真搞糊涂了——” ([3], p. 722)。群山发起并见证了诸多变换，但其自身是恒常不变的，凡尘中的人与事的变化，如瑞普的衰老，政权的更迭都不影响群山的恒常，瑞普的故事终结后，群山依旧伫立长存，这也使得继“Rip”的情节索引消失后，尾注部分再度出现了“mountain”。而另一方面“mountain”与“Rip”在开篇和尾注的分离也说明“mountain”的存在使故事形成了一个闭环，故事自“群山”始又到“群山”终。从现实层面来看，群山环绕着瑞普所在的村庄；在叙事层面，“mountain”建构的圆形闭环也将村庄中发生的所有故事都囊括其中，从而完成开篇群山“威凌四周的乡村”这一断言。

4. 主题词分析

本文在生成高频词表操作的基础上，收集了1778~1817年间创作的14部英语文学作品作为参照语料库，导入AntConc-Tool Preferences-Keyword List-Reference Corpus-Use Raw Files-Add Directory中，利用AntConc的Keyword List得出文本的主题词表，见表3，其中Keyness——主题性，可以反映主题词对原文主题建构的贡献程度。

Table 3. Keyword list generated by AntConc

表 3. AntConc 生成的主题词表

Rank	Freq	Keyness	Keyword	Rank	Freq	Keyness	Keyword
1	64	+722.45	Rip	17	7	+50.76	neighborhood
2	35	+385.68	Van	18	5	+48	Nicholas
3	30	+338.37	Winkle	19	4	+45.09	henpecked
4	20	+109.78	village	20	4	+45.09	ninepins
5	9	+101.46	Dutch	21	4	+45.09	ravine
6	9	+91.04	Wolf	22	9	+43.93	hat
7	15	+79.97	mountain	23	13	+43.35	mountains
8	9	+79.58	Dame	24	5	+40.11	liquor
9	7	+78.91	Hudson	25	4	+40.09	keg
10	10	+78.2	gun	26	8	+39.1	tree
11	31	+76.83	old	27	4	+37.46	beard
12	6	+67.63	flagon	28	4	+37.46	neighbors
13	11	+59.04	dog	29	3	+33.81	Congress
14	6	+58.65	cocked	30	3	+33.81	Knickerbocker
15	5	+56.36	Kaatskill	31	4	+32.76	Indians
16	5	+56.36	Vedder	32	11	+27.56	strange

根据主题词表，可以得出，对《瑞普·凡·温克尔》主题建构具有较大影响的元素涉及：1、人物：Rip Van Winkle, Dame Van Winkle, Nicholas Vedder, a dog named Wolf, neighborhood, Knickerbocker, Indians；2、地点/空间：Dutch village, Kaatskill mountains, Hudson, ravine, tree；3、事件：gun, flagon/liquor/keg, cocked hat, ninepins, Congress, beard；4、基调词：old, henpecked, strange。结合这些主题词进行文本细读，可以推定《瑞普·凡·温克尔》的主题大致包括：

Continued

cluster 6	人物形象、情节、批评、生态解读、短篇小说、背景、自然环境
cluster 7	偏见、典型形象、女性言说、枷锁、温克尔太太、男权、话语
cluster 8	元功能、功能文体学、叙事学、帝国主义叙事、男权主义
cluster 9	乡村、华盛顿·欧文、历史语境、反清教主题、文体分析、美国早期历史
cluster 10	历史书写、政治焦虑、民族神话、焦虑、记忆、身份
cluster 11	《桃花源记并诗》、乌托邦、逃遁、隐逸
cluster 12	哥特式主题、哥特式人物、哥特式情节、哥特式环境
cluster 13	《贾奉雉》、巴赫金、跨时空叙事
cluster 14	仙乡淹留叙事模式、狄俄尼索斯精神、神话原型批评
cluster 15	对比、美国浪漫主义文学、衬托
cluster 16	文化身份、文化体制、美国精神
cluster 17	协和观、新历史主义、遁世文学
cluster 18	《睡谷传说》、模仿、讽刺
cluster 19	主题、互文性理论
cluster 20	父权社会
cluster 21	集体无意识
cluster 22	政治反讽

以上聚类反映出目前学界对《瑞普·凡·温克尔》的研究情况，对比前文依据主题词表推定的5大主题可知两者大范围是重叠的，这说明基于主题词表的主题推定具有一定科学性和准确度，同时，值得注意的是，前文推定出的5大主题包含“印第安”，而该主题在国内现有研究的主题聚类中却并未出现，表明该主题尚未引起学界的普遍重视。有趣的是，《外国文学评论》2021年04期刚好刊登一篇名为“瑞普为何沉睡：《瑞普·凡·温克尔》中的政治焦虑”，文中就集中探讨了“印第安”这一主题，文章指出在刻画瑞普的人物形象时，欧文特别为他的个性中融入了印第安人的特质，而瑞普游离在白人主流价值观之外的情形便是此种特质的外化。在《附言》中，欧文一反常态地用略带贬义的“红人”(red man)指代印第安人，似乎是刻意地在二元对立的种族语境下暗示瑞普是“白人”与“红人”的矛盾统一体。因此，瑞普的边缘化无疑间接指向印第安人在美洲大陆被边缘化的现实境遇，瑞普的沉睡也由此衍生出印第安人的沉睡之意[4]。可见，通过远读(distant reading)确实可以有效辅助研究者发现文本的潜在主题。下文拟结合文本细读，对主题词表中的部分主题词展开细致分析：

4.1. “Gun” “Wolf” ——时间指涉与情节完善

“枪”和“狼”成为主题词的原因在于其对时间的指涉及其对后文重要的推动作用。从“干净的、擦足了油的枪”变得“包着一层铁锈，扳机已经脱落，枪托也蛀空”，二十年时光的流逝被作者刻意凝聚在“枪”的物理和化学变化之中。我国梁朝任昉《述异记》中也曾记载一个类似《瑞普·凡·温克尔》的故事：“信安郡石室山，晋时王质伐木至，见童子数人棋而歌，质因听之。童子以一物与质，如枣核，质含之，不觉饥。俄顷，童子谓曰：‘何不去？’质起视，斧柯尽烂。既归，无复时人。”在《瑞普·凡·温克尔》中，时光流逝同样也是借助入山者手头工具的物理变化展现的。二者都不约而同选择将时间寓于空间之中，借助空间所有者的状态变化映衬出岁月的磨蚀。除了指涉时间，“枪”也在空间层面推动了

生成《瑞普·凡·温克尔》的语义网络图，最后利用 MDS (Multidimensional scaling of geodesic distances) 算法进行聚类，聚类结果如图 6。

该聚类直接反映了故事的重要情节，值得注意的是，文本聚类的结果显示“尼古拉斯”也是该故事的重要一环，且单独被安置为一个聚类，而这一角色在前人的相关研究中都没有提及，可以说“尼古拉斯”是一个被读者遗忘的重要角色。回到文本发现，尼古拉斯·维德尔既是村长，又是客店的老板。而这个客店是“一个由村中的圣贤、哲学家和其他有空闲的人组成的永久俱乐部” ([3], p. 717)， “这个秘密政治会议(junto)里的意见，完全控制在他的手里” ([3], p. 717)。欧文在故事里特意将这个客店称为“junto”，而“junto”取自本杰明·富兰克林于 1727 年在费城创立的 Junto Club (共读社)，其原义是指为一个共同目标而集结的一群人(a group of people assembled for a common purpose)，意在通过成员间的分享和讨论，提升彼此的英语水平 and 人文学修养。富兰克林是一位坚定的清教代表，他注重理性，讲究实效，“有用”是他处世哲学中的首要原则。在美国那个启蒙运动的年代，游手好闲被视为一种罪过，因为这是一种背弃社会责任和个人责任的行为。瑞普奉行及时行乐的生活哲学，与富兰克林“辛勤劳作，推迟享乐”的劝诫背道而驰。富兰克林创建的共读社最初的 12 名成员是商人和工匠，他们在周五晚上聚会，讨论道德、政治或自然哲学问题。俱乐部持续了 38 年。

巧合的是，共读社的 12 名成员中恰好就有一位核心成员就叫“Nicholas”。同时，结合文本中对永久俱乐部和俱乐部成员的描述，可以明显看出作者的反讽和消极态度。反观欧文对“Nicholas”结局的设定：“尼古拉斯·维德尔！唔，他死了十八年了！本来在教堂的墓地里，他坟上还有一块木碑，上面刻着他一生的事迹，现在那块木碑也烂掉了，什么也没留下” ([3], p. 726)。欧文刻意安排了“Nicholas”的死亡，同时进一步加码，让其死后连“木碑也烂掉了，什么也没留下”。杨金才认为欧文对新英格兰北方佬及他们的清教传统持敌视态度。这种反清教思想在他的另一则故事《睡谷传奇》(*The Legend of Sleepy Hollow*)中也有同样的表现[5]。早在欧文的《纽约外史》一书中欧文就对清教徒做了相当细致的描绘并把他们写成了一群入侵者或歹徒。可以说故事中的“Nicholas”是当时清教的代名词，欧文对其命运的刻意安排实则暴露出反清教真实意旨，是作者精心安排的清教死亡。从这一角度来看，也就不难解释为何“Nicholas”为何对主题建构至关重要了。

4.3. Indians——失落的北美印第安

结合 Concordance Plot 的情节定位(见图 7)，可以发现“Indians”竟然也贯穿了故事首尾，同时在文本的边缘位置——尾注部分——密集出现，侧面暗示了故事主题与印第安边缘化的相关性。

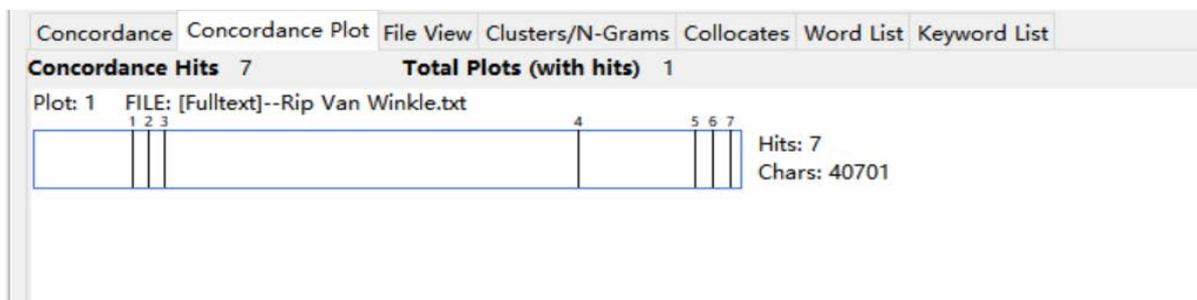


Figure 7. Concordance plot of “Indians”

图 7. “Indians”情节索引

反观“Indians”的第一次出现——“他(瑞普)参加他们的游戏，给他们做玩具，教他们放风筝和弹石子，并给他们讲关于鬼怪、巫婆和印第安人的长篇故事” ([5], p. 715)。这段描写透露出瑞普对印第安种

族和文化的熟识。学者金烁锋认为“瑞普和印第安人之间存在极大的相似性。似乎是刻意地在二元对立的种族语境下暗示瑞普是‘白人’与‘红人’的矛盾统一体”。在某种意义上，瑞普代表了入侵山林的白人，同时其自身也有着深刻的印第安烙印。

“瑞普一觉醒来时看到的是那头象征新型美利坚合众国的老鹰而不是原先受神仙召唤时见到的那只乌鸦”[5]，而乌鸦刚好是北美印第安人的重要图腾。乌鸦神话在北美洲印第安部落神话中大量流传，分布范围遍布北美洲版图的各个区域，无论是美洲大陆西南地区的热带地区还是北部的寒冷地区，都有乌鸦神话的存在。这些乌鸦神话无一不传达着印第安先民对乌鸦信仰的信奉。可以说，印第安部落先民视乌鸦为无所不能的神圣的鸟。印第安部落先民对于乌鸦怀有朴素而单纯的崇拜之情，也深刻地影响了他们的文化风俗，突出表现在他们隆重的节庆仪式中。在印第安地区至今还保留着狂热又盛大的敬奉乌鸦的太阳舞[6]。在《瑞普·凡·温克尔》中有两处对乌鸦的描写，第一处出现在瑞普被山中幽灵召唤之时：“他向四外一望，连个人影子也没有，只见一只乌鸦孤零零地振翼掠过山头”([3], p. 717)。值得注意的是，在《附言》中，作者讲述了一段印第安神话，指出卡斯基尔山是一个充满寓言的地区，这里的妖精女王可以随意让好运或厄运降临山谷。那些大神或者妖精居住在卡斯基尔山最荒凉的幽深处，捉弄在此处迷路的印第安猎人，甚至有猎手在这儿丢掉了性命，因此卡斯基尔山始终让印第安人感到虔诚的敬畏。这段看似松散的印第安神话的插入恰好暗示了乌鸦是印第安神灵现身的预兆，如《附言》所述，如果有人惹怒了卡斯基尔山中的妖精，就必定会有“灾难降临”，这无疑是对入侵印第安人领土的白人发出的警示。而瑞普苏醒后回到“通向圆剧场的那块地方”，“回答他的却是一群闲鸦的呱呱声，它们在高高的天空中，绕着一株倒挂在阳光照耀着的悬崖上的枯树盘旋。它们因为在很高的地方，觉得很安全，似乎正在向下面望着这个可怜人，嘲笑他的狼狈境况”([3], p. 719)。在这里，闲鸦的叫声就像印第安神灵的嘲笑声。同时乌鸦这个意象还埋藏了另一条线索——乌鸦与鹰的对立。通过文本细读可以发现，两次出场的乌鸦的实际状态都不好——第一次出场是“孤零零”，而第二处出场的乌鸦则是绕着枯枝盘旋才觉得安全。乌鸦代表着印第安种族和文化，而鹰代表着新生的美利坚合众国。文中对乌鸦的消极描写其实而暗示了印第安现实的糟糕处境。从美国历史来看，独立革命结束后，美国政府对待印第安人的政策发生了改变，印第安人由战时的潜在盟友和战争伙伴变成被征服的对象。在同一片土地上的印第安人并不是美国公民，也不受联邦政府保护。此后美国政府加快了西进运动的步伐，也进一步加剧了对印第安人的掠夺和伤害。欧文一直十分关注印第安人的境遇，对他们抱有深切的同情。在同收录于《见闻札记》的《印第安人性格的特点》一文中，欧文高度赞扬了印第安人坚忍淳朴的性格，哀叹他们被战争剥夺了土地[4]。也正因此在该故事中作者有意识地将“Indians”放逐在文本的边缘——尾注部分，同时刻意让具有印第安烙印的瑞普沉睡以刻意避开独立战争，让瑞普至少可以在战争中保持一种中立态度。

5. 结语

数字人文技术支持的远读可以发掘隐藏在文本中的潜在线索，结合文本细读和文学批评则可更加深入地探究文学作品的主题建构和内在价值。高频词“old”与荷兰早期移民的幽灵形象反复勾连，形成怀旧的复调，暗示故事逃遁与历史书写的主题。另一高频词“mountain”是鬼魅的祖地，也是自然之母的子宫，其在故事中的出场将所有情节囊括其中，形成自“山”始(开篇)至“山”终(附言)的情节闭环，“mountain”的恒常与历史的沧桑形成对比，体现出故事的自然神思想。主题词“gun”和“wolf”指涉时间，透露政治的跌宕及身份认同的失真。主题词“Nicholas”是共读社的代名词，其墓碑的朽坏体现了欧文对新英格兰清教传统的敌视态度。主题词“Indians”在故事中以“乌鸦”为化身，乌鸦的消极描写暗示了印第安现实的糟糕处境，欧文刻意将“Indians”放逐在文本的边缘——附言部分，也透露出作者对印第安失落的同情。本文只是对部分主题词的尝试解读，其余的主题词如“ninepins”，笔者认为“九柱戏”在最开

始本是宗教驱邪的仪式，后在 18 世纪演变为美国流行的娱乐游戏，故事中“九柱戏”造成的雷鸣般的隆隆声，一方面暗示了美利坚当时如火如荼的驱巫运动，另一方面也喻指了瑞普入睡后无法听到的独立战争的炮火声。限于篇幅，笔者未在正文部分予以展开，后续将进一步寻找例证，尝试对相关主题词进行解读。同时，上述的主题词表也留存了部分问题尚待解决，如主题词“cocked hat”和“beard”与故事的主题构建又有何关系，它们为何能成为主题词，同时又该如何解读？这或许需要后续研究者的结合数字人文方法进一步分析和发掘。总的来说，本文论证了结合数字人文技术辅助文本细读从而实现文本潜在主题挖掘的可行性，是文学跨学科研究的有效尝试与实践。

基金项目

本文受湖南省大学生创新创业训练计划项目(项目批准号：S202210542074)资助，使用该项目自建的大型英汉历时语料库。

参考文献

- [1] 胡开宝, 杨枫. 基于语料库的文学研究: 内涵与意义[J]. 浙江大学学报(人文社会科学版), 2019, 49(5): 143-156.
- [2] 董洪川, 潘琳琳. 数字人文与外国文学研究范式转换[J]. 西南民族大学学报(人文社科版), 2018, 39(9): 174-179.
- [3] Irving, W. (1961) *The Sketch Book of Geoffrey Crayon, Gent.* New American Library Inc., New York.
- [4] 金烁锋. 瑞普为何沉睡: 《瑞普·凡·温克尔》中的政治焦虑[J]. 外国文学评论, 2021(4): 26-45.
- [5] 杨金才. 从《瑞普·凡·温克尔》看华盛顿·欧文的历史文本意识[J]. 解放军外国语学院学报, 2001, 24(6): 75-79.
- [6] 郭美辰. 从乌鸦神话看北美洲与东北亚民间文化关联[D]: [硕士学位论文]. 长春: 东北师范大学, 2017.