

基于语料库软件AntConc对《麦琪的礼物》的研究

肖玉青

南京理工大学外国语学院, 江苏 南京

收稿日期: 2024年3月25日; 录用日期: 2024年5月20日; 发布日期: 2024年5月31日

摘要

语料库语言学是20世纪80年代兴起的一门研究文本检索、抽样、分析和统计的学科。本文以语料库语言学为基础,并综合利用语料库分析研究工具AntConc的“Word List”、“Concordance”、“Concordance Plot”、“File View”等功能,就美国知名小说家欧·亨利(O. Henry)的作品《麦琪的礼物》(*The Gift of the Magi*)的叙事视角、语言特色、人物形象、角色关系以及文本主题五个方面展开研究,以期为文学鉴赏开辟新的思路。

关键词

语料库语言学, AntConc, 《麦琪的礼物》

An Analysis of *The Gift of the Magi* Based on the Corpus Software AntConc

Yuqing Xiao

School of Foreign Studies, Nanjing University of Science and Technology, Nanjing Jiangsu

Received: Mar. 25th, 2024; accepted: May 20th, 2024; published: May 31st, 2024

Abstract

Corpus linguistics is a discipline that emerged in the 1980s and is devoted to text retrieval, sampling, analysis and statistics. Based on corpus linguistics, by making comprehensive use of “Word List”, “Concordance”, “Concordance Plot”, “File View” and other functions of corpus software AntConc, this paper analyzes the narrative perspective, language features, character images, character relations and text themes of the well-known American novelist O. Henry’s work *The Gift of the Magi*, so as to open up a new idea for literary appreciation.

Keywords

Corpus Linguistics, AntConc, *The Gift of the Magi*

Copyright © 2024 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

随着科技的飞速发展,计算机越来越广泛被应用在语言和人工智能方面。语料库语言学,是在 20 世纪 80 年代提出的一种新兴学科。从那开始,文本语料库的发展快速而完整,并越来越表现出了其实际意义。

欧·亨利是美国现代短篇小说之父,他发表了许多作品。《麦琪的礼物》是欧·亨利最著名的作品之一。虽然是短篇小说,但作者很好地展现了美国下层人民的生活和身处那个时代的一对年轻夫妇的美好爱情。欧·亨利作为短篇小说大师,其作品的语言风格值得深入探讨。不少学者对《麦琪的礼物》进行了研究,但大多局限于文学理论分析。基于此,本文尝试用语料库研究软件 AntConc 的词表(Word List)、索引(Concordance)、索引定位(Concordance Plot)、文件查看(File View)和搭配(Collocates)等功能,从叙事视角、语言特征、人物形象、人物关系和文本主题这几个方面来解读《麦琪的礼物》。本文的研究问题如下:

1) 语料库软件 AntConc 可否被用于分析文学作品《麦琪的礼物》?

2) 如果语料库软件 AntConc 可被用于分析文学作品《麦琪的礼物》,那这篇小说的叙事视角、语言特色、角色关系、人物形象及文本主题是什么?

2. 文献综述

AntConc 是日本早稻田大学 Laurence Anthony 博士实验室基于微软 Windows 操作系统开发的多平台语料库分析工具包,界面简洁、操作方便[1]。至 2024 年 3 月,在国内 CNKI 数据库中输入“AntConc”进行精准检索,可获得 2008~2024 年相关文献 197 篇。近十几年来, AntConc 软件对相关文献的研究显示,语料库选择呈现多元化趋势。如谢程程的《基于 Antconcl 的中美百强大学校训之比较研究》,就以中美百强大学校训为数据源,构建了一个封闭的语料库。在 AntConc 分析的基础上,研究了中美两国大学校训在词源、构词、内容、文体等语言维度上的差异[2]。杨懿靖以语料库软件 AntConc 为工具建立了主题词表,来分析《人民日报》(海外版)中“全球文明倡议”的相关报道[3]。本文以英文文本欧·亨利的短篇小说《麦琪的礼物》为研究对象,借助 AntConc 软件对《麦琪的礼物》这篇小说进行分析。

很多学者对《麦琪的礼物》进行了研究,但大多集中在文学领域。如牛玫的《探析英文短篇小说〈麦琪的礼物〉中小人物的大情》就是通过研究人物情感来分析这篇小说[4]。相比之下,从语言学的角度对《麦琪的礼物》的研究却有些不足。因此,本文以语料库语言学为基础,尝试从客观性的角度解读《麦琪的礼物》。

3. 研究方法

3.1. 研究语料

《麦琪的礼物》讲述的是一对生活艰苦的年轻夫妇忍痛割爱互赠圣诞礼物的故事。本文使用语料库

分析软件 AntConc 3.5.9 对《麦琪的礼物》的英文文本进行分析。这篇小说的字词总数是 1952。

3.2. 研究工具

本文以语料库软件 AntConc 为研究工具，其作用是利用语料库对文本进行分析，它包含词表、索引、索引定位、文件查看、搭配等功能。其中，词表可以罗列出文本中涉及的所有单词，并计算每个单词在文本中的出现次数，反映其词频信息；索引可以列出所有包含搜索词的句子片段，有利于句子的上下文分析；索引定位可以反映检索项在文本中的分布信息；文件查看可以进一步帮助我们了解文本内容；搭配是一种共现信息，可以提供搜索词的句法和语义信息[5]。

4. 检索结果及分析

4.1. 用词表和索引分析叙事视角和语言特色

词表和索引能反映出小说的叙事视角和语言特征。词表可以使我们对小说的主要人物、叙事视角和语言特征有一个大致了解。索引可以提供检索词在语料库中的上下文信息。为了分析这部小说，我们使用词表功能对《麦琪的礼物》中出现频率最高的 30 个单词进行检索，结果如表 1 所示。

Table 1. Top 30 high-frequency words

表 1. 前 30 个高频词

Rank	Freq.	Word	Rank	Freq.	Word	Rank	Freq.	Word
1	101	the	11	24	was	21	17	he
2	74	and	12	23	you	22	17	that
3	65	a	13	22	had	23	14	hair
4	44	of	14	20	for	24	11	as
5	39	to	15	19	with	25	11	me
6	33	it	16	18	della	26	11	they
7	33	she	17	18	his	27	11	were
8	27	her	18	18	i	28	10	said
9	26	jim	19	18	on	29	9	all
10	24	in	20	17	at	30	9	but

首先，从表 1 可以看出，文中的人称代词主要有“it”、“she”、“her”、“you”、“I”和“he”。其中，“she”出现了 33 次，排名第 7；“he”出现 17 次，排名 21。因此，可看出《麦琪的礼物》主要采用的是第三人称的叙事视角，这种视角的优势在于可以不受时间和空间的束缚，比较客观地描述事实、反映内容。从表 1 中人物的排序也可以清楚地看到，“Jim”出现的次数最多，为 26 次，“Della”紧随其后，为 18 次。这说明 Jim 和 Della 是小说中的重要人物。其次，“you”出现了 23 次，在第 12 位，“I”出现了 18 次，在第 18 位。结合索引功能发现，这两个人称代词多出现在人物的对话中，其中“said”出现了 10 次，说明小说中有较多的对话描写。通过这些对话，我们可以进一步推断出故事人物和情节。此外，从表 1 可以看出，“was”是出现频率最高的动词，出现了 24 次。除“was”外，还有很多其它动词的过去式，如“had”、“were”和“said”。因此，可以得出结论，《麦琪的礼物》是以第三人称的视角来讲述过去发生的故事。

4.2. 用词表、索引、文件视图和搭配功能分析角色关系和人物形象

根据表 1，可以发现文本中出现了两个人物，Jim 和 Della，所以可以猜测他们是这部小说的主角。然后利用索引定位功能分别搜索“Jim”和“Della”，结果如图 1 和图 2 所示。从两幅图中可以看出，这两个名字的位置几乎贯穿了整个文本，这表明故事是围绕着他们展开的。在情节的开头和结尾，我们可以发现，在“Jim”的部分，“Della”出现较少；相反，在“Della”的部分，“Jim”出现的较少，说明这两个部分分别描述了男女主角的形象和行为。除这两部分外，他们出现的地方大多是重合的，可推测他们之间存在密切联系，可能是恋人关系。接着，我们结合文件查看功能得出他们确实是一对夫妇，这也印证了我们的猜想。

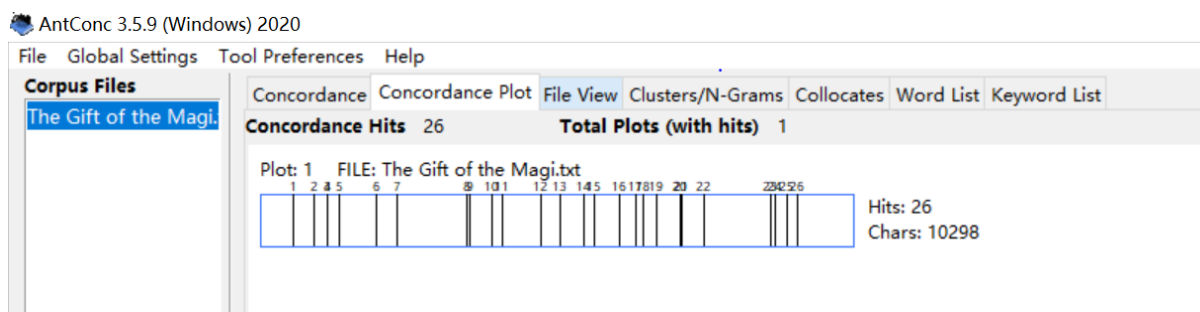


Figure 1. Index location diagram for “Jim”

图 1. “Jim” 的索引定位图

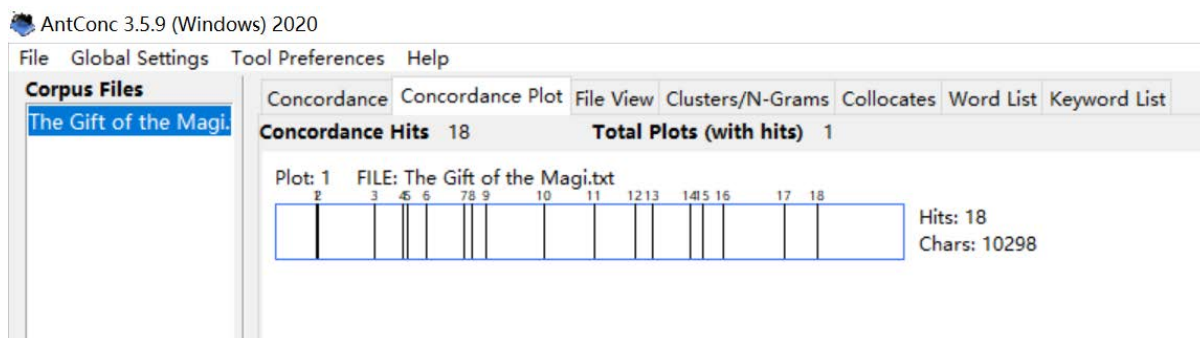


Figure 2. Index location diagram of “Della”

图 2. “Della” 的索引定位图

语料库研究的第一步虽然只能检索出语言的符号形式，但是由于它能够同时呈现出这些形式的语境共现和频率分布，这些符号就可能逐渐变成具有各种意义和功能的信息[6]。在上述表 1 中，“she”一共出现了 33 次。通过索引功能可得出，“she”指的是 Della 这个角色。因此，“Della”在小说中被提到了 51 次。结合搭配功能，与“Della”搭配的名词主要有“hair”、“present”、“penny”，动词有“whispers”、“cried”、“save”、“cut”、“faltered”等。结合文件查看功能，可得知 Della 有一头可以引以为荣的秀发。她一直在存钱，想给 Jim 买一份像样的圣诞礼物，但因钱不够而感到难过。最后她卖掉了自己的头发，给 Jim 买了一条昂贵的表链。

同样，“he”在这部小说中出现了 17 次。通过索引功能，我们可以发现“he”指的是 Jim。由于“Jim”出现了 26 次，所以“Jim”总共被提到了 43 次。根据搭配功能可看出，与之匹配的名词主要有“watch”、“present”，动词“need”、“entered”、“looks”、“dropped down”等，形容词“thin”、“serious”

等。结合“文件视图”，我们得知 Jim 努力工作，随着圣诞节的临近，为了给 Della 一份令人印象深刻的礼物，他不惜卖掉了三代怀表。

4.3. 用索引和文本查看功能分析文本主题

本节拟利用语料库软件 AntConc 对《麦琪的礼物》的主题进行分析。根据前一节的分析，我们可以推断出这部小说的主题是爱情。在本研究中，使用 AntConc 的索引功能依次搜索“love”和“like”，“love”的搜索结果为 2 次，“like”的搜索结果为 8 次。因为“like”还可以作介词，结合“索引”，看出“like”作为动词时仅出现两次。通过文件查看功能，我们得知包含 love 的两句显示出了 Della 对 Jim 的爱意之深沉，以至于她愿意剪掉自己心爱的头发，为 Jim 买一件珍贵的礼物。包含“like”的两句表达出 Della 害怕失去美丽的长发后 Jim 不再爱她，而 Jim 表示无论如何，他都不会减少对 Della 的爱，这两句反映出 Della 和 Jim 对彼此的深厚情感。根据以上分析，我们可以感觉到歌颂爱情是这部小说最直观的主题。

其次，宣扬人性之美也是小说的主题之一。据以上分析，以及 Della 和 Jim 为了给对方买礼物而放弃自己宝贵的东西，我们也可以推断出他们的经济状况差强人意。结合文本视图功能，我们得知《麦琪的礼物》这个故事反映了当时社会贫富差距悬殊，Jim 和 Della 所处社会底层阶级，生活极其艰苦。在当时的时代背景下，底层群体往往代表着贪婪、肮脏、犯罪等等，但欧·亨利笔下的底层群体，特别是 Jim 和 Della，被赋予了甘愿牺牲、积极向善的形象。因此，小说也颂扬了这样一种观点：婚姻和人生的最大快乐并不在于财富的多少，而在于源自人性向善、人性之美的相互理解 and 自我满足[7]。

5. 结论

通过以上分析，可得出 AntConc 可以用来研究文学作品。本文借助 AntConc 软件，综合运用词表、索引、索引定位、文件查看、搭配等功能，分析得出短篇小说《麦琪的礼物》是以第三人称的叙述视角，讲述了主人公 Jim 和 Della 的角色关系及形象；并得出歌颂爱情和宣扬人性美是这篇小说的主题。

运用语料库来分析文本确实能让我们更加客观地认识文学作品，但也存在一定的局限性。首先，仅通过搜索几个单词来推测人物形象和主题是不够的，这部作品还有很多主题，用语料库软件并不能完全将这些主题展现出来；其次，一些检索词的选择也比较主观。因此，通过语料库软件 AntConc 分析得到的结果可以作为参考，但是如果觉得通过这些词不足以精确地分析人物形象或主题的话，可以结合更多词或者结合文本细读来达到更深的分析目的。

参考文献

- [1] 赵建军, 徐祖熹. 基于语料库工具 AntConc 对《故乡》的研究[J]. 湖北第二师范学院学报, 2019, 36(1): 23-27.
- [2] 谢程程. 基于 AntConc 的中美百强大学校训之比较研究[J]. 软件导刊(教育技术), 2016, 15(10): 90-91.
- [3] 杨懿靖. 《人民日报》(海外版)“全球文明倡议”报道的话语建构研究——基于 AntConc 语料库分析[J]. 中国传媒科技, 2024(1): 64-67.
- [4] 牛玫. 探析英文短篇小说《麦琪的礼物》中小人物的大情[J]. 戏剧之家, 2021(20): 191-192.
- [5] 张萌. 基于语料库工具 AntConc 的《面纱》研究[J]. 现代交际, 2021(19): 91-93.
- [6] 高博. 基于语料库的埃兹拉·庞德《诗经》英译研究[J]. 外国语言文学, 2013, 30(3): 173-180+216.
- [7] 肖丹. 《麦琪的礼物》的主题解析与表现艺术[J]. 语文建设, 2016(35): 77-78.