

基于语料库软件AntConc对小说《边城》的分析

刘明慧

辽宁师范大学文学院, 辽宁 大连

收稿日期: 2025年8月12日; 录用日期: 2025年9月1日; 发布日期: 2025年9月15日

摘要

运用语料库软件AntConc4.2.4W对沈从文中篇小说《边城》进行量化研究, 采用语料库的方法对叙事特征、主要人物、语言特色、主题等进行分析, 直观地展现小说的艺术性。研究发现, 高频词“了”“翠翠”“祖父”等揭示了小说的叙事时态、核心人物及对话驱动的叙事特点; 通过人物在文本中的分布及搭配词分析, 清晰呈现了翠翠的纯真形象、祖父的慈爱特质、二老的性格特征及他们之间的情感关联; 语言层面, 高频比喻词、单纯词及数字的运用凸显了作品质朴灵动的语言风格; 核心词“爱”及相关文本分析则深化了对小说人性美与自然美主题的理解。本研究不仅为《边城》的文学阐释提供了实证支持, 发掘了文本潜在意义, 也验证了语料库方法在解析中国文学作品显性与隐性内容方面的有效性和可行性。

关键词

AntConc, 《边城》, 写作艺术

Analysis of *Border Town* Based on Corpus Software AntConc

Minghui Liu

School of Chinese Language and Literature, Liaoning Normal University, Dalian Liaoning

Received: Aug. 12th, 2025; accepted: Sep. 1st, 2025; published: Sep. 15th, 2025

Abstract

The corpus software AntConc4.2.4W is used to quantitatively study Shen Congwen's novella *Border Town*, and the corpus method is used to analyze the narrative characteristics, main characters, language characteristics, themes, etc., and intuitively show the artistry of the novel. It is found that the high-frequency words "Liao", "Cuicui", "grandfather", etc. reveal the narrative tense, core

文章引用: 刘明慧. 基于语料库软件 AntConc 对小说《边城》的分析[J]. 现代语言学, 2025, 13(9): 478-484.

DOI: 10.12677/ml.2025.1391002

characters and dialogue-driven narrative characteristics of the novel. Through the distribution of characters in the text and the analysis of collocations, Cuicui's innocent image, grandfather's loving traits, the personality characteristics of the two elders and the emotional relationship between them are clearly presented. At the language level, the use of high-frequency metaphors, simple words and numbers highlights the simple and flexible language style of the work; the core word "love" and related text analysis deepen the understanding of the novel's themes of human beauty and natural beauty. This study not only provides empirical support for the literary interpretation of *Border Town*, explores the potential meaning of the text, but also verifies the effectiveness and feasibility of the corpus method in analyzing the explicit and implicit content of Chinese literary works.

Keywords

AntConc, *Border Town*, Writing Art

Copyright © 2025 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

自 20 世纪 80 年代起, 计算机技术的持续进步为语料库的发展提供了强大动力, 使其应用范围不断拓展。语料库分析方法的应用领域经历了显著扩展: 最初仅局限于计算语言学范畴, 随后逐渐被理论语言学、描述语言学及应用语言学研究广泛采用, 如今已进一步延伸至认知语言学和翻译学领域。这一发展历程充分彰显了语料库分析方法在语言学研究中的关键地位[1]。在语料库语言学蓬勃发展的大背景下, 语料库文体学这一概念应运而生。作为一门交叉学科, 语料库文体学借助语料库语言学的研究方法, 将文学文本分析与文学批评有机融合。随着语料库语言学研究的不深入, 语料库文体学也日益受到学界的广泛关注和重视[2]。语料库文体学主要借助 WordSmith、AntConc 等语料分析工具, 通过对文学文本(尤其是小说和译作)进行词汇统计、语境分析、搭配模式、词串提取、叙事推进及关键词识别等多维度研究, 实现定量与定性相结合的分析。这种方法能够深入探究文学作品的人物刻画、叙事结构、核心主题及作者风格等特征, 为传统文体学研究提供了重要的方法论补充[3]。

近年来, 国内学界开始运用语料库分析工具涉足文学文本研究, 相关探索主要集中在以下几个方向: 一是围绕词频统计与主题词分类展开分析, 像王雁针对《傲慢与偏见》开展的语料库检索研究便属此类; 二是聚焦词语前后搭配情况进行探讨, 例如杨建玫对《警察与赞美诗》的语料库检索分析; 三是以具体词语(如人名、地名等)作为检索对象进行剖析, 周艳丽与张发祥关于《德伯家的苔丝》的语料库检索分析即是典型案例; 四是对文本中反复出现的短语(即词簇或词丛)加以研究, 以此揭示文学作品的文体特征, 陈婵所作的“爱丽丝·门罗小说中的词簇特征及其功能分析——一项基于语料库的文体学研究”就体现了这一方向; 五是对国内文学作品进行语料库检索分析, 比如陈婵、程乐通过提取莫言小说《蛙》的主题词, 从主题、故事情节及人物形象塑造等维度展开了较为深入的探讨[4]。

《边城》是沈从文的代表作, 在中国近代文学史上具有独特的地位。《边城》以 20 世纪 30 年代川湘交界的边城小镇茶峒为背景, 以船家少女翠翠与船总家儿子傩送的情感纠葛为主线, 在讲述这段凄美纯粹的爱情故事的同时, 也向世人展示了湘西独有的民俗风情。

作为中国现代文学的重要代表人物, 沈从文的文学作品一直受到学术界的持续关注。以《边城》为例, 中国知网数据库显示相关研究成果数量可观, 仅标题含该关键词的文献就超过千篇。现有研究主要

从多个维度展开，包括生态美学解读、跨文化传播研究、人物形象分析(如翠翠角色塑造)以及建筑艺术特征探讨等。然而值得注意的是，这些研究大多沿袭传统的文学批评范式，缺乏方法论层面的创新。

在此研究背景下，本研究拟引入语料库语言学的研究方法，对《边城》进行定量与定性相结合文本分析。这一尝试具有双重研究价值：一方面可为既有的文学阐释提供数据支撑，另一方面有望发掘文本中潜藏的新意义。同时，本研究也将验证语料库方法在解析中国文学作品的显性与隐性特征方面的适用性与有效性。

2. 运用语料库软件 AntConc 对小说《边城》进行直观而深入的分析

(一) 运用词表 “Word List” 分析叙事特征

通过 Word List 对《边城》文本中高频词汇进行检索，绘制成下表 1。

Table 1. Statistical table of high-frequency words in the text of *Border Town*
表 1. 《边城》文本中高频词汇统计表

Type	Rank	Freq	Type	Rank	Freq
的	1	1241	到	25	170
了	2	922	来	26	169
翠翠	3	523	上	27	162
在	4	374	一个	28	156
人	5	339	时	28	156
你	6	322	船夫	30	148
着	7	320	便	31	145
说	8	312	二老	32	144
不	9	284	看	33	137
祖父	10	280	只	34	133
去	11	267	还	35	126
我	12	265	个	36	124
一	13	263	同	37	122
把	14	225	得	38	120
他	15	224	两	39	119
船	16	222	她	40	117
就	17	213	却	41	116
这	18	211	什么	42	106
那	19	210	事	43	95
也	20	191	为	44	90
有	20	191	想	45	89
老	20	191	要	45	89
是	23	174	过	47	88
又	24	172	很	48	83

通过对表 1 中观察，“了”字出现 922 次，位于第 2 位，可以推测出《边城》是叙述过去发生的事

情,“翠翠”出现了 523 次,位于第 3 位,可以推测出她是文章的主人公,以女性角色为主要描写对象。“祖父”出现了 280 次,位于第 10 位,可以推测出他也是小说中的主要人物。“说”出现 312 次,位于第 8 位,是《边城》中出现频率最高的动词,可以推测出《边城》中有大量人物对话存在,可以通过人物的对话推测故事情节发展和人物性格特征[5]。

在人称代词方面,“你”“我”均出现在高频词表中,又通过 Cluster 功能检索发现,“你”常和“你的”“你说”“你在”、“你来”搭配,出现在句子对话中;“我”常和“我就”、“我要”、“我也”、“我听”、“我说”搭配,大部分存在于人物的对话之中,由此不难看出,该小说并非采用第一人称或第二人称来展开叙述,而是以第三人称进行叙述的。

(二) 运用词表“Word List”、索引定位“Concordance plot”和搭配“Collocates”以及文件查看“File View”功能分析主要人物及其关系

依据 Word 的词频检测结果,“翠翠”一词共出现 523 次,是该文本中出现频次最高的人物。在索引定位功能“Concordance plot”中输入“翠翠”后,可获取其在情节中的分布图,如下图 1 所示:

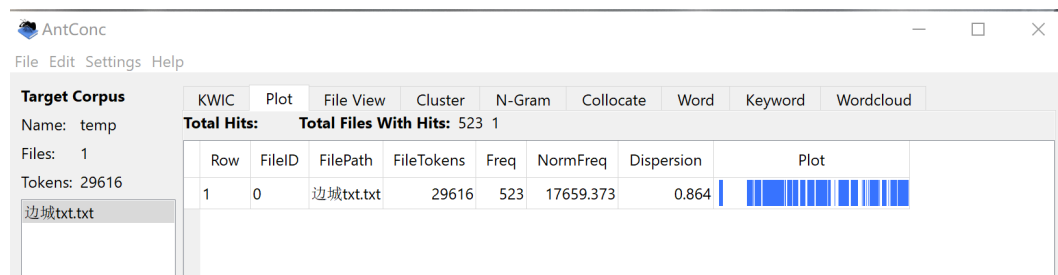


Figure 1. Plot distribution map of “Cuicui” in the novel *Border Town*

图 1. “翠翠”在小说《边城》中的情节分布图

从“翠翠”的情节分布图能够发现,“翠翠”贯穿于小说的全过程,在整篇作品中均频繁出现,是小说的主人公。使用搭配功能“Collocates”里检索人物“翠翠”,发现和她搭配的词有形容词天真善良、温柔清纯,我们可以推测小说中的翠翠是个纯洁美丽的山村少女,根据“File View”可知,翠翠终日与清澈透明的小溪,与青山绿水的大自然为伴。并且发现她和祖父关联最多,说明对祖父十分依赖,哪怕是去仅有一里路程的城里看热闹,也不和爷爷分开,同爷爷一起撑船一起歇息,不愿意让爷爷一个人独处。白塔、渡船、黄狗,祖孙俩相依为命。

根据 Word List 的词频检测结果,结合表 1。可知小说中“祖父”这一称谓出现 280 次,而“爷爷”的频次为 78 次。通过文本文件检索功能(File View)可以确认,这两个称谓均指向同一人物角色。

在索引定位“Concordance plot”里输入“祖父”“爷爷”,得到“祖父”“爷爷”在小说《边城》中的情节分布图,如下图 2、图 3 所示:

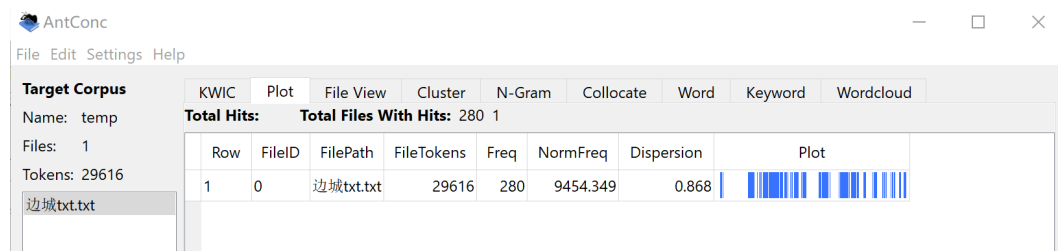


Figure 2. Plot distribution map of “Grandfather” in the novel *Border Town*

图 2. “祖父”在小说《边城》中的情节分布图

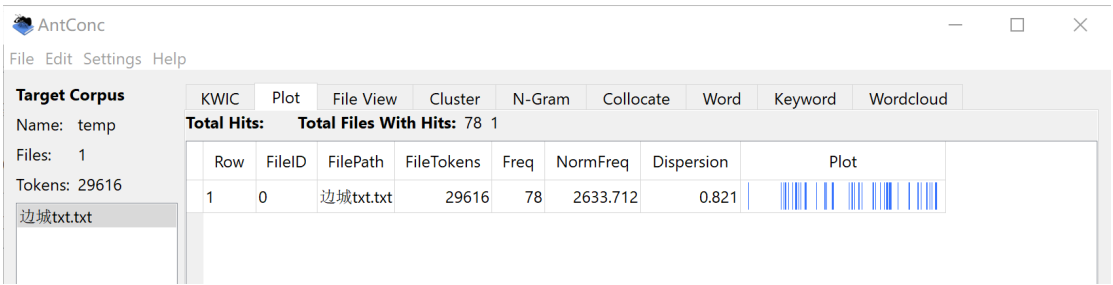


Figure 3. Plot distribution map of “Yeye” in the novel *Border Town*
图 3. “爷爷”在小说《边城》中的情节分布图

从“祖父”“爷爷”的情节分布图可以看出，“祖父”也几乎贯穿于整篇小说之中，是文中出现次数第二多的人物。通过 File View 功能，爷爷时常与翠翠交流，生活中对她关怀备至，感情上也能体谅她的心事，还为翠翠的美丽天真而感到骄傲与欣慰，并且在翠翠遇到困难的时候，希望能够利用自己微薄的力量，维护翠翠纯真的爱情，爷爷因为自己女儿婚姻的悲剧，在心中留下了深深的伤痛，他非常害怕女儿的悲剧在自己孙女翠翠身上重演。借助一连串的语言刻画，既凸显了爷爷的淳朴敦厚，也充分展现出他对翠翠的疼爱。

在索引定位“Concordance plot”里输入“二老”，得到“二老”在小说《边城》中的情节分布图，如下图 4 所示：

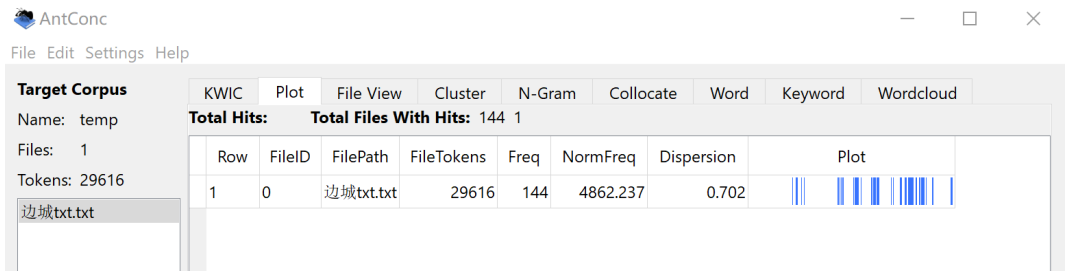


Figure 4. Plot distribution map of “Erlao” in the novel *Border Town*
图 4. “二老”在小说《边城》中的情节分布图

从“二老”的情节分布图能够看出，“二老”几乎贯穿了整部小说，是其中的关键人物。不过，“二老”的情节分布图前期存在一定空白，与图 1 中“翠翠”的情节分布图相比，这表明“二老”的出场时间相对较晚。运用搭配功能“Collocates”对“二老”这一人物进行检索后，发现与之搭配的形容词包括惊讶、羡慕等，通过 File View 功能，发现他对翠翠一见钟情，当他为翠翠唱歌的时候，二老倾诉了自己浓浓的爱意，用自己的歌声打动了翠翠；当王团总以碾坊作为陪嫁时，二老不为所动，仍然追求自己的真爱；当得知哥哥闯滩而死的时候，二老心怀愧疚，最终选择了出走。可以看出他拥有善良的品性，具备勇于追求的精神，对金钱与地位持轻视态度，并且有着极强的责任感。

(三) 运用词表“Word List”、索引定位“Concordance plot”和搭配“Collocates”以及文件查看“File View”功能分析工具分析文章的语言特色

(1) 《边城》中的比喻修辞手法

利用“Concordance”功能检索“如”“一样”，出现的频率分别是 24 次、14 次，这说明《边城》使用了大量修辞手法，使得整部作品生动、自然而又活灵活现。沈从文在《边城》的第一章刻画翠翠时，运用了大量比喻手法。比如“翠翠在风日里长养着，把皮肤变得黑黑的，触目为青山绿水，一对眸子清明如水晶”“人又那么乖，和山头黄鹿一样”，前者以水晶喻指翠翠的眼眸，后者将翠翠比作山巅的黄

鹿。这些比喻的运用，把湘西的山水与人物紧密联结在一起，让翠翠的形象更加鲜活丰满，也由此展现出湘西民众特有的精神风貌。

(2) 《边城》中单纯词语的运用

利用“Concordance”功能检索“只”“到”的频率分别为133次和177次，并且查看表一高频词汇表发现在《边城》中，沈从文常运用单纯词来描绘湘西景致、讲述翠翠的爱情故事。例如“人若越过小溪小山走去，只一里路就到茶峒城边”中的“人”“只”“越”“到”，以及“月光极其柔和，溪面浮着一层薄薄白雾，这时节对溪若有人唱歌，隔溪应和，实在太美丽了”中的“极”“浮”“层”“薄”等词语，不仅体现了沈从文简约的语言风格，更能以精炼的文字承载丰富的内涵，为读者留下广阔的想象空间，同时增添了小说的典雅韵味，使其更具吸引力。

(3) 数字的运用

利用“Concordance”功能检索数字“一”，出现的频率是263次，在《边城》里，沈从文多次运用数字，既展现出文字的简洁之美，又能让读者获得视觉上的美感。中国古典名著《红楼梦》也常以数字的运用增强作品的艺术性，读来更令人赏心悦目。作为知名作家，沈从文在《边城》中同样频频借助数字营造视觉美感。比如文中“由四川过湖南去，靠东有一条官路，这官路将近湘西边境，到了一个名叫‘茶峒’的小山城时，有一小溪，溪边有座白色小塔，塔下住了一户单独的人家”这段文字，连续使用数字“一”，不仅让行文简洁明了，还提升了读者的理解度，以简练的方式呈现出湘西文化中独特的乡土风情[6]。

(四) 运用核心词表“Keyword”、索引定位“Concordance plot”以及文件查看“File View”等功能分析《边城》的情节发展与主题

要领会一部小说的核心意涵，就得把握其叙事线索与贯穿全篇的情节。前文借助Word List、Concordance Plot、Concordance及Collates等功能，对小说的叙事角度、语言风格、人物形象及相互关系做了简要剖析，虽对小说内容有了大致认识，但对文本脉络的掌握还不够全面细致。翠翠作为小说的主人公，全文围绕她展开，根据情节分布图可知，中间有一部分没有关于翠翠的描写，通过“File View”可知，那部分主要在叙述湘西的美丽环境，《边城》开篇就将读者带入湘西那片绚丽的景致中：茶峒这座城依山傍水而建，靠山的一侧，城墙宛如一条长蛇，沿着山势蜿蜒而上。靠水的一侧则在城外河边留出空地设置码头，供小小的蓬船停靠。白河下游到辰州与沅水汇合后，河水便略显浑浊，带有山间泉水的气息。如果逆流而上，那些三丈五丈深的水潭都清澈见底。在日光的映照下，潭底的小白石子、带花纹的玛瑙石子都清晰可见。水中的游鱼来来往往，仿佛悬浮在空气中。两岸多是高山，山中盛产可用来造纸的细竹，常年呈现出浓郁的翠绿色，格外醒目。靠近水边的人家大多坐落在桃杏花丛中，春天的时候只要留意就会发现，凡是有桃花的地方必定有人家，有人家的地方必定可以买到酒。夏天，阳光下晾晒的紫花布衣裤光彩夺目，如同标识人家所在的旗帜。到了秋冬季节，那些建在悬崖上、滨水处的房屋，都清晰可见。这些房屋有着黄泥的墙壁、乌黑的瓦片，位置总是那么适宜，与周围环境极为协调，给人带来十分愉悦的直观感受。作者时而浓墨重彩，时而轻描淡写，笔下的景致美不胜收，让人仿佛亲身置身于湘西边地的山光水色之中。他以充满诗意的语言，精心构建出一个人与自然万物和谐统一、密不可分的世界，生动地展现了边城的环境之美。

对人性情感的深度探索始终是文学创作的核心命题。沈从文在《边城》中通过独特的艺术表现，系统呈现了其关于美与爱的哲学思考，这部作品堪称作家创作生涯中最具人文关怀的代表作。通过语料库方法对“爱”这一核心词汇进行系统检索与分析，有助于研究者准确把握文本的叙事重点和主题内涵[7]。在《边城》勾勒的湘西世界中，人与人之间涌动着至纯至无私的情感。首先是老船夫与翠翠的祖孙情深：老船夫一心牵挂孙女的幸福，为她奔波操劳，乃至夜不能寐、心力交瘁。其次是大老与二老的兄弟之谊，

通过 Cluster 功能可发现,“爱”与两人均有搭配;结合“File View”可知,兄弟二人同时倾心于翠翠,不可避免地展开了爱情角逐——这场“走车路”与“走马路”的竞争,是坦荡磊落的君子之争,即便如此,他们的兄弟情依旧真挚,甚至愿为这份情谊将爱情暂且搁置。比如二老将公平视为竞争的前提,提议两人轮流唱歌,由翠翠自主抉择。此外,还有翠翠与二老之间纯粹的爱恋,“File View”呈现了他们相爱过程中的诸多温馨片段,像小说第十四节中,翠翠在梦中因二老的动人歌声,灵魂仿佛被托起,轻轻飘荡,掠过白塔、菜园、小船,又飞至溪对岸的高崖采摘虎耳草。这一情节生动印证了,那种纯净而朦胧的异性之爱,能让年轻男女沉浸在难以言喻的幸福感中[8]。

在文学作品的客观分析方面,语料库检索研究凭借其客观的文本分析视角展现出显著优势。它所提供的鉴赏结论可信度较高,若将这种研究方式与传统的主观鉴赏相结合,便能为文学鉴赏领域开辟新的路径。

3. 结语

将语料库分析手段应用于文学文体学研究,为文学文本分析带来了全新视角,如同为感性的文学文本解读赋予了理性的支撑,有助于提升解读文学作品的精准度。本文借助 Antconc 这一语料库工具,从叙事视角、人物、语言、主题等维度对《边城》展开分析,这种以理性方法研究富有情感的文学作品,或许缺少传统文学欣赏所具备的感染力。不过,该方法严谨客观,结构层次分明,分析结果清晰易懂。它不仅不会损害文学欣赏的韵味,反而能弥补欣赏过程中可能存在的主观性、片面性等问题。当然,我们必须承认,用语料库工具分析文学作品无法替代从文学自身角度进行的赏析,但它确实为我们理解文学作品提供了有力依据[9]。验证了基于语料库的研究方法在揭示中国文学作品中显性与隐性内容时,具有实际应用的有效性和可行性。

参考文献

- [1] 卫乃兴. 语料库语言学的方法论及相关理念[J]. 外语研究, 2009(5): 36-42.
- [2] 卢卫中, 夏云. 语料库文体学: 文学文体学研究的新途径[J]. 外国语(上海外国语大学学报), 2010, 33(1): 47-53.
- [3] 姜晓艳. 基于语料库的《简·爱》语言特点及主题词表征分析[J]. 江苏科技大学学报(社会科学版), 2016, 16(2): 77-84.
- [4] 陈婵, 程乐. 莫言作品《蛙》的语料库检索分析[J]. 浙江工商大学学报, 2014(5): 26-34.
- [5] 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.
- [6] 白永红. 湘西文化《边城》中的语言特色[J]. 宿州教育学院学报, 2019, 22(1): 34-36.
- [7] 吕会红, 刘韵盈. 利用语料库软件 AntConc 对《Hamlet》的多维度分析[J]. 信息与电脑(理论版), 2020, 32(23): 89-91.
- [8] 刘伟安, 陈俐. 《边城》爱的主题新论[J]. 集美大学学报(哲学社会科学版), 2015, 18(4): 84-89.
- [9] 郑文扬. 利用语料库软件 AntConc 对《追风筝的人》的多维度分析[J]. 中国多媒体与网络教学学报(上旬刊), 2020(4): 247-248.