

# 基于语料库的《小妇人》三词词簇分析

张 艳

上海海事大学外国语学院, 上海

收稿日期: 2023年6月5日; 录用日期: 2023年7月13日; 发布日期: 2023年7月24日

## 摘 要

近年来, 利用语料库及语料库分析软件分析文学作品成为一种新的研究理念与方法。本研究正是基于这种方法, 运用语料库分析软件AntConc对《小妇人》这部小说中的高频三词词簇进行挖掘并对“I do not”以及“I am afraid”这两个三次词簇进行详细分析。研究发现: 三次词簇即其搭配可以展示主人公的心理及性格, 还可以进一步深挖作者创作文本背后的深层含义。

## 关键词

《小妇人》, 三词词簇, 语料库文体学

# A Corpus-Based Analysis of Lexical Bundles with Three Words in *Little Women*

Yan Zhang

School of Foreign Languages, Shanghai Maritime University, Shanghai

Received: Jun. 5<sup>th</sup>, 2023; accepted: Jul. 13<sup>th</sup>, 2023; published: Jul. 24<sup>th</sup>, 2023

## Abstract

In recent years, using corpus and corpus analysis software to analyze literary works has become a new research concept and method. This study is based on this method, using corpus analysis software AntConc to excavate high-frequency lexical bundles with three words in the novel *Little Women* and conduct a detailed analysis of the two lexical bundles with three words: “I do not” and “I am afraid”. The study has found that lexical bundles with three words and their collocations can show the protagonists’ psychology and personality and can further explore the deep meaning behind the author’s creative text.

## Keywords

### *Little Women*, Lexical Bundles with Three Words, Corpus Stylistics

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

经典长篇小说《小妇人》是美国女作家路易莎·梅·奥尔科特的代表作，被称作“女孩成长的《圣经》”。这是一部以女主人公乔为原型的半自传体小说。它讲述了马奇家的四个女孩在美国内战期间由女孩逐渐成长为女人的故事。随着语料库的不断发展，利用语料库及语料库分析软件分析文学作品成为一种新的研究理念与方法。作为一种新兴的研究领域，语料库文体学融合了语料库语言学与文学文体学两个学科的研究范式，用计算工具进行人文研究[1]。关于《小妇人》的研究方兴未艾，学界现有研究多从女性主义等传统文学研究视角切入，鲜有从数字人文视域切入，即用语料库文体学的方式分析该本小说。因此，为弥补研究空白，本研究另辟蹊径，基于文学文本语料库，运用语料库检索软件 AntConc 获取小说中的核心三词词簇并分析其的语用功能，旨在探析《小妇人》情节发展，主题意义及作者路易莎·梅·奥尔科特的叙事风格。

## 2. 词簇与语料库文体学

根据胡开宝，语料库文体学是“采用语料库方法，在分析文学文本的语言特征并进行相关数据统计的基础上，深入探讨文学文本特征、作家风格以及文学与社会的关系等问题[2]”。薛冰(2020)运用 DocuScope 对《论语》的英译本进行了研究，研究发现 DocuScope 等数字化人文研究工具为文学翻译提供了新的研究视角与研究方法[3]。凤群(2014)基于语料库对弗吉尼亚·伍尔夫意识流小说《达罗卫夫人》进行文体学分析，反映了伍尔夫的写作特点及写作风格，体现了语料库文体学在分析文学文本方面强大的功能[4]。孙萌(2022)等依托语料库文体学视阈，以太宰治《人间失格》为例，对日本近现代文学的异质性特征进行研究。在大数据时代背景下，语料库文体学为文学文本提供了新的解读方式，为文学文本的分析注入新的源泉[5]。Moss (2014)选取亨利·詹姆斯的小说章节自建语料库，探析亨利·詹姆斯小说的句法功能[6]。Egbert (2012)基于语料库文体学，通过量化分析了十位作家共一百部作品，探索了十九世纪小说的文本特点[7]。总而言之，关于语料库文体学的研究方兴未艾，其将定量研究与定性研究相结合，从全局的视角对文学文本的翻译、文本意义以及作者的写作风格进行探究，是对传统文学研究的有益补充。

从语料库研究的角度来看，Biber 等人将词簇定义为“在给定语域中最常见的单词组合(word combinations that recur most commonly in a given register) [8]”。词簇也称词块、词串等。一般来说，语言学中，三词词簇与四次词簇的意义更大。王薇(2022)等运用语料库分析软件 AntConc 的词簇等功能分析了怀特黑德的小说《第一区》，为语料库文体学的研究提供了新的探索与启示[9]。陈婵(2014)对爱丽丝·门罗小说中的词簇特征及其功能进行分析，研究结果表明部分核心词簇对小说的情节发展、角色塑造以及主题分析起到了至关重要的作用，语料库文体学有助于发现以往传统文学定性研究中没有注意到的文本意义[10]。

### 3. 《小妇人》高频三词词簇分析

笔者在研究前运用语料库分析软件 AntConc 的 N-Gram 的功能对《小妇人》[11]中的词簇进行检索后发现：二词词簇出现频率过高，且存在意义不完整的现象，而四词词簇出现频率则过低，因此，本研究选取三词词簇作为研究对象，笔者运用 AntConc 检索出《小妇人》中使用最多的 8 个三词词簇(见表 1)，再通过 AntConc 的 KWIC (Key Words In Context)功能将这些三词词簇定位到小说的语境中进行分析，以期了解奥尔科特的写作风格和主要人物的性情。

**Table 1.** High-frequency lexical bundles with three words in *Little Women*

**表 1.** 《小妇人》中高频三词词簇

词簇	频数
I do not	220
I can not	102
In spite of	72
I did not	60
I am not	48
As well as	48
I will not	45
I am afraid	40

从表 1 不难发现，从人物本身的角度出发，即以第一人称的方式直观地展示人物的主要活动的高频词簇共有 6 个(I do not、I can not、I did not、I am not、I will not、I am afraid)，说明小说以第一人称视角叙述。且 8 个高频三词词簇中，有 4 个三词词簇包含助动词(do, have, be)，这表明该小说中隐含了大量的动作描写，而情态动词(can、will)的出现则表明《小妇人》中有许多关于主人公心理活动的描述。表 1 中共有 5 个三词词簇是否定形式的(I do not, I cannot, I did not, I am not, I will not)，说明小说的很多对话以及动作具有否定含义，这也是《小妇人》的一个行文特征[12]。

词簇 I do not 的详细分析

从图 1 不难发现，小说《小妇人》中词簇 “I do not” 的分布相对较为平均。根据 AntConc 的 KWIC 功能，词簇 I do not 与 know 搭配 30 次，与 think 搭配 24 次，与 believe 搭配 24 次，与 see 搭配 16 次，与 care 搭配 12 次，这些搭配均用于描述人物心理活动。因此可知，在《小妇人》中，“I do not” 被用来展示主人公的心理状态以及在对话中的情绪。以下是具体的两则例子。

Row	FileID	FilePath	FileTokens	Freq	NormFreq	Dispersion	Plot
1	0	Little Women.txt	191178	221	1155.991	0.920	

**Figure 1.** KWIC of lexical bundles “I do not”

**图 1.** 词簇 “I do not” 的索引定位

例 1: “You, you are, you’re a great deal too good for me, and I’m so grateful to you, and so proud and fond of you, **I don’t know** why I can’t love you as you want me to. I’ve tried, but I can’t change the feeling, and it would be a lie to say I do when I don’t.”

例 1 是乔所说的一段话。乔(Jo)是《小妇人》的女主人公。乔是个率真的女孩,具有男孩子气,与传统观念所要求的女性形象格格不入,她敢爱敢恨却又善良正义。乔一直走的都是独立自主的女性路线,她认为女性应该有自己的思想,远大的理想,女性也有自己的天赋。以上的例子便是乔在拒绝富二代劳里(Laurie)追求时的对话。笔者认为乔对劳里是有男女之情的,但由于劳里的贵族身份,乔嫁给劳里就得做贵族太太,很难再做自己。与其说乔拒绝的是劳里,不如说乔拒绝的是贵族的身份与不愿意画地为牢、束缚自我的生活。“I don't think”展现了女主人公乔在回应求爱的对话中的心理状态与沮丧的心境,展现了主人公的性情。小说中的乔是一个典型的女权主义者,她甚至用自己的行动来抵制社会对女性的偏见,她从不觉得婚姻是女性的使命与任务,这与作者路易莎·梅·奥尔科特的女性主义立场是一致的。

例 2: “Then I'll go without. **I don't care** what people say!” cried Jo, taking up her book.

例 2 是女主人公乔与姐姐梅格(Meg)的对话。对话的情境是两姐妹要赴约参加加德纳夫人举办(Mrs. Gardiner)的舞会,姐姐梅格要求乔一定要戴手套,因为不戴手套就不能跳舞。在这种情形下,乔说出了“那就不带好了,我才不管别人怎么说呢”。这句话中, I do not 与 care 搭配,表明乔并不在乎别人看法,展现了乔的人物性格。作者所刻画的乔一直都是不在意世俗眼光的人,她讨厌受到别人的影响,渴望自由自在的生活。通过阅读小说的前半部分,笔者发现乔对上流社会有抵触情绪,她的这句话中的“I don't care”也体现了她对这种上流社会约定俗成的规则以及上流贵族评价的不屑。

以上两则例子均是三词词簇“I don't”搭配表示心理状态的词,展现了主人公的人物形象、性格。也让读者看到了《小妇人》这部小说的主旨:女性成长与女权意识。

词簇 I am afraid 的详细分析

根据图 2 不难发现,“I am afraid”这个三词词簇在小说前半部分的使用频率比小说后半部分高。使用 AntConc 的 KWIC 对“I am afraid”的语境进行定位,不难发现“I am afraid”后面所接的词语或语句多半具有否定含义。《小妇人》是一部关于女性成长的小说,讲述了马奇家的四个女儿由女孩成长为坚韧且有力量女性的故事。大女儿梅格美丽动人,一开始的她会认为自己的礼服在朋友里显得很寒酸且向往上流的贵族世界。但后来,梅格清醒了,不屈服于所谓的金钱论,在选择爱情时选择遵从自己的内心;二女儿乔独立自主,热爱写作。她是个有点假小子形象的女性,她直率坦荡,但不可避免地脾气暴躁,但后来,乔学会了反省并改变暴躁的脾气。三女儿贝丝是典型的“贤妻良母”式女孩,她单纯善良,默默付出不求回报,后来因身体原因早殇;小女儿艾米热爱绘画,希望成为优秀的画家。艾米随着成长越来越知道自己要的是什么,她目标明确且敢于争取。马奇家的四个女儿在不断地成长,她们内心越来越丰盈,越来越坚定且无所畏惧,因此小说地后半段“I am afraid”这个三词词簇出现频率较之前半段有明显降低。这也体现了作者路易莎·梅·奥尔科特的创作意图,即注重表现女性意识,强调女性成长,宣扬女性的美好品质。关于“I am afraid”在小说中出现的具体例子如下。

Row	FileID	FilePath	FileTokens	Freq	NormFreq	Dispersion	Plot
1	0	Little Women.txt	191178	40	209.229	0.838	

Figure 2. KWIC of lexical bundles “I am afraid”

图 2. 词簇 “I am afraid” 的索引定位

例 3: “You don't know, you can't guess how bad it is! It seems as if I could do anything when I'm in a passion. I get so savage, I could hurt anyone and enjoy it. **I'm afraid** I shall do something dreadful some day, and spoil my life, and make everybody hate me. Oh, Mother, help me, do help me!”

在《小妇人》第八章中，乔在艾米意外落水后因自己的脾气暴躁差点造成遗憾而深深自责，母亲告诉乔：“每个人都有自己的诱惑，有些甚至比这种大得多，我们常常要用一生的时间来征服它们。”母亲的同情和信任显然是很有说服力的。后来乔逐渐变得稳重，能够更好地控制自己的脾气。恐惧未来让乔获得成长，让乔学会抵制诱惑。

例 4: Meg colored, but a mischievous fancy to tease the girls made her reply demurely, “You are very kind, but I’m afraid he won’t come.”

例 4 的情形是梅格踏足名利场，参加舞会时的场景。梅格惊叹于莫法特家华丽的房屋与主人的优雅，惊羡于朋友安妮·莫法(Annie Moffat)的富有，并沮丧于自己的拮据。舞会时劳里给梅格送来鲜花，这让梅格的虚荣心得到了很大的满足。以下的对话就是舞会上梅格的朋友揶揄她想邀请劳里来参加舞会时，劳里逗弄以及回答朋友的话。“I’m afraid”后面的“he”指代的就是劳里。从这个片段不难发现，此时的梅格抱怨不满于自己生活，她想要踏上上流社会，她期待得到上流社会人们的赞美，更想找到一个有钱的丈夫。她觉得朋友的善意都是对她贫穷的怜悯。显然，梅格这短时间的经历让她感觉到不快乐，她回家向妈妈哭诉这几天的经历，妈妈告诉梅格“我绝不希望你们觉得金钱是最重要的或者是唯一的奋斗目标”“我宁愿看到你们成为拥有快乐、幸福美满的穷人的妻子，也不愿你们成为没有自尊、没有安宁的皇后”。在母亲的引导下，梅格获得了成长，在最终择偶时她选择了爱情，嫁给了一位没地位但值得爱的男人。因此我们也可以看到梅格的成长，摆脱虚荣、接纳自己也接纳爱人。

以上两则例子是乔和梅格的话语，在话语中，她们均使用了“I’m afraid”这个三词词簇，这些话语也来源于小说的前半部分。通过通读小说后，笔者见证了她们的成长与越来越坚定的内心。因此在小说前半部分出现更多的三词词簇“I’m afraid”。

#### 4. 结语

本研究运用语料库与语料库分析软件 AntConc 剖析文学文本《小妇人》，探析小说中高频三词词簇及其语用功能。研究发现通过分析三次词簇即其搭配可以展示主人公的心理及性格，还可以进一步深挖作者创作文本背后的深层含义。笔者通过对三词词簇“I do not”以及“I am afraid”进行详细分析后发现女性成长是贯穿于《小妇人》行文的一条主线，也该小说的主旨。马奇家的四位女孩在母亲的教导下，不管在成长中多迷茫，犯了多少错误，有过多少荒唐的想法，最终都会敢于作出改变，遵从内心的自由做出自己的选择。这也是《小妇人》作者路易莎·梅·奥尔科特想要传递给读者的理念。

#### 参考文献

- [1] Gold, M.K. (2014) Digital Humanities. In: Ryan, M., Emerson, L. and Robertson, B.J., Eds., *The Johns Hopkins Guide to Digital Media*, Johns Hopkins University Press, Baltimore, 143-148.
- [2] 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.
- [3] 薛冰, 向明友. 数字人文视域下《论语》英译本的修辞对比研究[J]. 外语研究, 2020, 37(4): 83-93.
- [4] 凤群. 基于语料库的意识流小说《达罗卫夫人》文体学分析[J]. 山东外语教学, 2014(1): 43-47.
- [5] 孙萌, 刘晓芳. 语料库文体学视阈下日本近现代文学的异质性特征研究——以太宰治《人间失格》为例[J]. 解放军外国语学院学报, 2022, 45(4): 152-159.
- [6] Moss, L. (2014) Corpus Stylistics and Henry James’s Syntax. Ph.D. Thesis, University College London, London.
- [7] Egbert, J. (2012) Style in Nineteenth Century Fiction: A Multi-dimensional Analysis. *Scientific Study of Literature*, 2, 167-198. <https://doi.org/10.1075/ssol.2.2.01egb>
- [8] Biber, D., Johansson, S., Leech, G., Conrad, S., et al. (1999) *Longman Grammar of Spoken and Written English*. Pearson Educational Limited, London.
- [9] 王薇, 杨兴隆. 怀特黑德小说《第一区》的语料库文体分析[J]. 天津外国语学院学报, 2022, 29(6): 93-107.

- [10] 陈婵. 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[J]. 解放军外国语学院学报, 2014, 37(3): 151-159.
- [11] 路易莎·梅·奥尔科特. 小妇人[M]. 盛世教育西方名著翻译委员会译. 上海: 上海世界图书出版公司, 2010.
- [12] 杨元媛, 宋军. 基于语料库词簇的文学作品人物心理分析——以《劝导》为例[J]. 商丘师范学院学报, 2015(7): 91-94.