断裂与回归：《笨花》城市化进程的文学计算分析

Rupture and Return: A Computational Literary Analysis of the Urbanization Process in *Clumsy Flower*

杨革新（**Yang Gexin**）　林　啸（**Lin Xiao**）

内容摘要：本研究运用文学计算批评方法，对铁凝长篇小说《笨花》进行数字人文分析，探究其所呈现的城市化进程特征。通过文本预处理、词频统计、LDA 主题建模、循环神经网络情感分析等计算方法，将小说按历史时期划分为四个城市化阶段进行量化分析。研究发现，《笨花》中的城市化进程呈现“孤立—萌芽—互动—回归”的非线性波动格局，战争因素构成城市化进程的关键断裂变量。情感分析显示，与城市化相关的文本整体呈现负面情感主导特征，且负面程度随城市化推进而加深。本研究为理解 20 世纪上半叶中国乡村城市化的复杂性提供了以数据为支撑的文学证据。

关键词：《笨花》；城市化；文学计算批评；循环神经网络；情感分析；LDA 主题建模

作者简介：杨革新，浙江大学外国语学院教授，主要从事英美文学、西方文论和文学伦理学批评研究；林啸，浙江大学外国语学院博士研究生，主要研究方向为英美文学、文学伦理学批评研究。本文为国家社科基金重大招标项目“当代西方伦理批评文献的整理、翻译与研究”【项目批号：19ZDA292】的阶段性成果。

Title: Rupture and Return: A Computational Literary Analysis of the Urbanization Process in *Clumsy Flower*

Abstract: This study employs methods of computational literary criticism to conduct a digital humanities analysis of Tie Ning’s novel, *Clumsy Flower*, examining the characteristics of the urbanization process it portrays. Through text preprocessing, word-frequency statistics, LDA topic modeling, and recurrent neural network–based sentiment analysis, the novel is partitioned by historical period into four stages of urbanization for quantitative analysis. The findings indicate that the urbanization process in *Clumsy Flower* exhibits a nonlinear, fluctuating trajectory of “isolation-emergence-interaction-return,” with war functioning as the key rupture variable. Sentiment analysis shows that text related to urbanization is dominated by negative

sentiment, and the degree of negativity intensifies as urbanization advances. This study provides data-supported literary evidence for understanding the complexity of rural urbanization in China during the first half of the twentieth century.

Keywords: *Clumsy Flower*; urbanization; computational literary criticism; recurrent neural network; sentiment analysis; LDA topic modeling

Authors: Yang Gexin is Professor at the School of International Studies, Zhejiang University (Hangzhou 310058, China). He is primarily engaged in British and American literature, Western literary theories and ethical literary criticism (Email: ygx80080@163.com). **Lin Xiao** is a Ph.D. student at the School of International Studies, Zhejiang University (Hangzhou 310058, China). His research focuses on British and American literature, ethical literary criticism (Email: 2895865392@qq.com).

一、引言

铁凝的长篇小说《笨花》以 20 世纪上半叶华北乡村为背景，通过笨花村的兴衰变迁，展现了中国传统乡土社会在现代化进程中的深刻转型。作为新世纪以来重要的乡土叙事作品，《笨花》不仅延续了中国乡土文学的书写传统，更以其独特的叙事视角呈现了城市化进程的复杂性与非线性特征。然而，既往研究多从传统文学批评角度解读作品的主题意蕴与艺术特色，较少运用计算方法对其城市化叙事进行系统的量化分析。

近年来，数字人文方法为文学研究提供了新的分析工具和研究范式。莫雷蒂（Moretti）提出的“远读”（distant reading）概念强调通过计算方法发现文本中的潜在模式。[1] 马修 • 乔克斯（Matthew L. Jockers）进一步发展了“宏观分析方法”，利用文本挖掘技术探索大规模文学文本的深层结构。[2] 本研究以循环神经网络为基本工具，综合运用文本预处理、词频分析、LDA 主题建模[3]、情感分析等数字人文方法，对《笨花》的城市化进程进行多维量化分析。研究旨在回答以下核心问题：《笨花》如何通过词汇选择和主题分布呈现城市化进程？不同城市化阶段的文本特征有何差异？作品对城市化持何种情感态度？通过回答上述问题，本研究试图为理解中国乡土文学的城市化叙事提供新的分析视角。

1　参见 Franco Moretti, *Distant Reading*, London, New York: Verso, 2013, 45-67.

2　参见 Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary History*, Urbana: University of Illinois Press, 2013, 23-28.

3　参见 David M. Blei, Andrew Y. Ng and Michael I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research* 3 (2003): 993-1022.

二、研究方法与数据处理

2.1 语料与阶段划分

本研究首先将《笨花》全文转换为UTF-8格式，确保文本编码的统一性。基于小说的叙事时间线与核心情节转折，结合历史背景与城乡发展特征，将全文划分为四个对应不同城市化阶段的文本片段：

第一阶段为“前城市化乡村主导期”（第1-14节），聚焦笨花村传统劳作、宗族生活等纯农村场景，体现传统乡土社会的封闭性特征。

第二阶段为“被动城市化萌芽期”（第15-32节），通过向喜的军旅生活、向文成开办西医诊所、引入《复活》等新书等情节，展现城市消费文化对乡村伦理的侵蚀以及现代教育与医疗资源向乡村的渗透。

第三阶段为“城乡互动及战时混乱期”（第33-55节），聚焦战争语境下的城乡重构，日军占领华北导致城乡秩序崩溃，笨花村人口流动加剧。

第四阶段为“战时断裂与战后乡土回归期”（第56-62节），揭示战时乡村城市化的临时性与非持续性特征。

去噪环节剔除标点符号、页码、注释等非叙事干扰信息。分词环节建立“笨花城市化自定义词典”，确保城市化相关词汇不被错误拆分。停用词过滤采用哈尔滨工业大学停用词表，并根据文本特征进行补充调整。

2.2 文本预处理

文本预处理采用“去噪—分词—停用词过滤”三步流程。

去噪：剔除页码标记（如“第X页”“PXX”）、注释等非叙事信息，保留核心叙事。

分词：加载“笨花城市化自定义词典”（一行一词），确保城市化相关复合词不被误切。

停用词：在哈尔滨工业大学停用词表基础上增补口头语与语义空洞词，提升统计有效性。

词元映射：手动构建“城市化关键词-词元”映射表，将语义近似词归并，便于跨词形统计（如“土地/耕地/官地”→“传统农耕生产”词元），见表1。映射表保存为“词元映射表.xlsx”。

表1 “城市化关键词—词元”映射表

词元	包含的城市化关键词
传统农耕生产	土地、耕地、官地、居连、居连后园子、山药蔓子、庄稼、牲口
农村传统行为	赶大集、插佛堂、走动儿幽会、祭祖、解手、喝号、庆祝会

2.3 特征工程与指标

词云与词频/词元频率：循环读取四个阶段的文本，统计关键词与词元

频率，结合 Voyant Tools 白名单生成阶段词云。

文体特征：

词汇密度（LD）：$LD=\frac{N_{城市化关键词}+N_{其他实义词}}{N_{总词数}}\times 100\%$；

平均句长：$MSL=\frac{N_{总词数}}{N_{句子数}}$；长词比例：$LWR=\frac{N_{长度\geq 6的城市化关键词}}{N_{城市化关键数}}\times 100\%$

TF-IDF 与趋势分析：用 jieba.analyse.extract_tags() 提取各阶段 Top-20 关键词（加载停用词与自定义词典），挑选代表性的城乡与战时要素绘制趋势。

LDA 主题建模：gensim 训练 LDA。参数：主题数 K=2（经调参避免重叠与权重塌陷）、迭代 =15、alpha=0.6、随机种子 =100。低频词过滤、字典与词袋构建遵循标准流程。

循环神经网络（RNN）情感赋值：

样本筛取：仅保留含至少一个城市化关键词的句子。

词典先验：加载BosonNLP与HowNet情感词典作弱先验对比与误差诊断。

模型结构：使用双向 GRU 分类器（BiGRU），以预训练中文词向量初始化嵌入层；为了稳健性，采用小批量微调，Sigmoid 激活输出在 [-1,1] 区间，经温度标定与分层抽样验证；基线参考预训练情感模型（Erlangshen-RoBERTa-110M-Sentiment）以交叉校准，最终以 RNN 输出为主分析对象。

汇总与可视化：阶段均值柱状图与阶段分布饼图对比情感强度与结构。

2.4 关系与相似性分析

共现网络：按句级窗口构建四阶段核心词共现矩阵与热图（颜色越深共现越强）。

文档相似性：将分词文本转为 TF-IDF 向量，计算余弦相似度，绘制相似度热力图。

对应分析（CA）：以阶段为行变量、关键词（Top-20）为列变量，输入为频次矩阵，prince 降维至二维，绘制阶段与关键词的双标图。

三、研究发现

3.1 城市化关键词的动态演变

通过使用 Voyant Tools 工具，分别上传已预处理完成的四个阶段文本，将之前编写的“笨花城市化自定义词典”输入到“White List”（白名单）中，以每行一个词的格式粘贴，即可得到每个阶段仅包含城市化关键词的词云按第一阶段至第四阶段顺序排列如下（见图 1-4）：

词云可视化分析显示，四个阶段的城市化关键词呈现明显的动态变化特征。

第一阶段体现“农耕生产 + 农村居住”的核心叙事格局。词云图 1 中乡

图 1

图 2

图 3

图 4

土词（窝棚、牲口、黄土）字体最大，与《关键词词频表》中该阶段“牲口 42 次、窝棚 53 次、黄土 15 次”的高频数据吻合，体现“农耕生产 + 农村居住”的核心叙事。城市词（皮鞋 14 次、政府 1 次）字体小，仅为“城市文化对农村的零星影响”（如皮鞋作为城市服饰符号），未改变乡土主导格局。

第二阶段呈现城乡词汇的结构性转变。词云图 2 中乡土词（土地 6 次、窝棚 16 次）字体缩小，对应《关键词词频表》中该阶段乡土词频率较前一阶段大幅下降（如牲口从 42 次降至 28 次），说明农村场景的核心地位松动；城市词“俱乐部（16 次）”“皮鞋（18 次）”字体增大：“俱乐部”是城市专属社交场所，“皮鞋”是城市物质符号，两者高频体现城市文化从物品（皮鞋）延伸到场所（俱乐部），“城里”高频则暗示城市作为空间概念开始被频繁提及，总体来说，城市元素从物质到组织层面渗透，乡土叙事占比下降。

第三阶段词云图 3 中“进城”出现爆发性增长，为 31 次（前两阶段最高仅 5 次），直接体现农村人向城市流动的核心叙事线，是城乡互动的标志性关键词，乡土词（“窝棚”55 次、“黄土”12 次）虽仍存在，但字体缩小，说明农村场景从主导变为战时依托。新出现的“八路军”反映战争对城市化进程的深刻影响。

第四阶段城市化元素几乎消失，词云图 4 中乡土词仅残留“牲口”，且无“城里”“皮鞋”“进城”等城乡流动词，体现城市化因战争影响而退潮，乡土叙事仅保留核心符号，无细节展开。四幅词云整体上体现了城市化从无到有，再到退潮的动态过程。

此外，通过结合《笨花》叙事逻辑，从各阶段 Top 20 关键词中，筛选出频率与权重随阶段显著变化、且与《笨花》中城市化进程相关的词，分成农村传统元素、城市关联元素、城乡空间关联、战时特殊元素四类别，导入 Python 中进行趋势折线图绘制，得出图 5 关键词演化趋势，进一步揭示了不同类型词汇的动态轨迹。

“棉花”属于农村传统符号，体现农村传统模式的动态变化。其在前城市化乡村主导期的“关键词相对占比”数值处于最高水平，反映此时《笨花》的叙事重心是农村社会，生活空间围绕农耕场景展开，体现了传统模式的绝对主导。在被动城市化萌芽期到城乡互动及战时混乱期，传统的生产需求随人口向城市流动而减少，其频率持续下降。这说明城市化进程直接冲击农村传统生产生活方式，传统元素的衰退是城乡资源流动的必然结果。在战时断

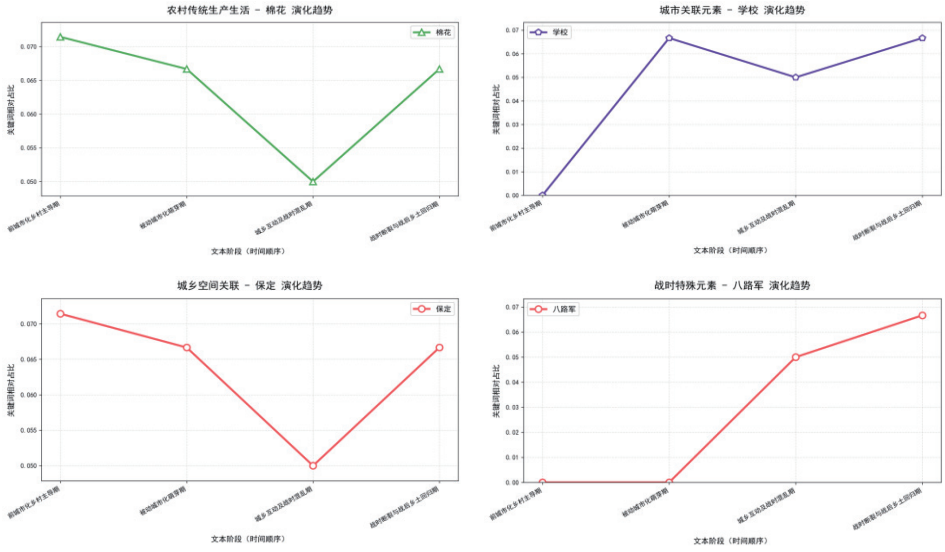

图 5　城市化关键词演化趋势

裂与战后乡土回归期，其频率略有回升，反映战时结束后乡土回归的叙事转向，说明《笨花》所描写的城市化并非单向消解农村或完全取代农村，而是受战时背景影响，出现传统元素的阶段性回调。

“学校”与“保定”属于城市导向符号，二者的趋势变化映射出城乡间元素渗透与空间连接的波动。在前城市化乡村主导期，学校作为城市文明的核心载体，频率接近 0.00，说明此时城乡处于低关联状态，城市文明元素未向农村渗透。保定作为城乡连接的核心城市空间，有较高的关键词占比，预示着城乡交互的到来。在被动城市化萌芽期“学校”的频率开始上升并达到最高，“保定”略有下降但依然较高，体现城市化萌芽的核心特征，即城市文明开始向农村渗透，城乡空间连接稳定。在城乡互动及战时混乱期，两类元素频率出现下降，凸显了战争对于城市化进程的严重破坏，“学校”因该阶段所开办的“夜校”而依然保有较高的关键词占比。在战时断裂与战后乡土回归期，两者的频率均出现回升，表明尽管战争导致了乡土叙事的回归，但城市化进程迅速再次走上正轨。

“八路军”是区别于农村或城市常规元素的断裂变量，其趋势独树一帜，体现战争对城市化进程的干扰。从前城市化乡村主导期到被动城市化萌芽期：“八路军”为零频率，无影响。说明这两个阶段的城乡关系围绕生产生活互动展开，无战时因素介入，城市化进程处于常规推进状态。在城乡互动及战时混乱期，其频率骤升，此时叙事重心从城乡经济和文明互动转向战时秩序。八路军的出现代表战争对城乡社会的全面介入，打破了农村——城乡互动——城市化深化的常规轨迹。在战时断裂与战后乡土回归期，其频率依然较高，体现了社会重新回归乡土本位，城市化进程开始常态恢复时，战

争对于城市化进程依然有着持久辐射。

这几张趋势图共同打破了“城市化为农村单向变为城市”的刻板认知，补充了战时特殊情境对于变迁轨迹的关键影响。

3.2 文体特征的阶段性

文体特征的量化主要通过词汇密度（LD）、平均句长、长词比例三个指标，量化城市化对《笨花》文本风格的影响，并用图表对比各阶段差异。其计算逻辑为 LD 越高，文本中城市化相关及实义信息越密集；句长越长，叙事越舒缓；句长越短，叙事越碎片化；长词比例越高，城市化叙事复杂度越高。通过调用前面使用的增补后的停用词表，并在 Python 中导入 Matplotlib 库，编写代码以批量计算各阶段问题指标并绘图，得到如下文体特征折线图（见图 6）：

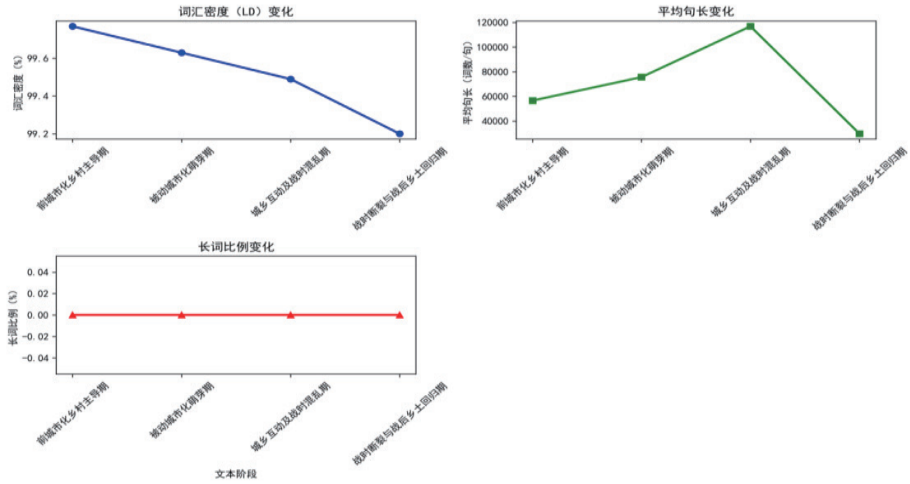

图 6 各阶段城市化相关文本研究体特征对比

文体特征量化分析显示，在词汇密度的方面，前城市化乡村主导期（阶段 1）词汇密度最高，文本中无城市化关键词，但乡土类实义词高度集中，一方面，叙事聚焦农耕生产与农村居住的具体细节，无冗余的非实义内容；另一方面，总词数中“停用词 / 无意义词”占比极低（因乡土叙事逻辑连贯，无需额外铺垫），导致“其他实义词数量 / 总词数”的比值达到峰值，最终使词汇密度成为全阶段最高。

被动城市化萌芽期（阶段 2）文本中新增城市化关键词（如皮鞋、俱乐部），但此类词多为“符号性城市元素”（非深度实义信息）；同时，乡土类实义词（土地、窝棚）的频率较阶段 1 降低，导致乡土实义词减少，而城市化关键词的“实义密度”（如“俱乐部”仅指代场所，无额外细节）低于乡土类实义词，最终“城市化关键词 + 乡土实义词”的总实义信息占比下降，词汇密度随之降低。

城乡互动及战时混乱期（阶段 3）的战时场景引入了较多非实义与重复性

信息，不计入实义词统计，打破了实义信息的集中性，导致词汇密度继续下降。战时断裂与战后乡土回归期（阶段 4）中，城市化关键词几乎消失，其他实义词数量大幅减少，导致词汇密度为全阶段最低。

这一趋势说明《笨花》的文本信息密度并非仅由城市化深化这一因素驱动，而是受城市化与战时事件的双重影响，实义信息随阶段推进持续降低，但就百分比来看并不特别显著，无明显断层。

在平均句长变化的方面，阶段 3 的平均句长最长，这主要是由战时场景导致的，战时场景虽紧张，但叙事聚焦同一空间或同一事件，避免了频繁场景切换导致的短句，最终使平均句长达到全阶段最高；阶段 2 同样包含了一定比例的战时场景，且叙事多为乡土场景与城市元素结合的衔接句，虽有场景切换，但逻辑连贯，无碎片化断句。阶段 1 几乎为纯乡土叙事，多为简洁的乡土动作描写，句子结构较为简单，因此平均句长低于阶段 2 与阶段 3，体现了乡土叙事的简洁性。长词比例四阶段延续 0 值，因《笨花》文本中城市化关键词多为 2-4 字，以简单符号为主，无复杂政策，无需额外修正。

3.3 主题分布与情感倾向

利用 LDA 主题建模，可从主题维度挖掘各阶段的核心叙事，通过主题词的更替印证关键词演化的结论，避免单一关键词分析的局限性。首先通过过滤低频词构建词典和词袋模型，再在 Python 的 Gensim 库输入为各阶段的分词文本，训练 LDA 模型，保持主题数适配城市化核心维度，并设置好能确保模型稳定的参数。在固定随机种子为 100 之后，项目针对“主题数”“迭代次数”“主题分布先验”三个变量通过控制变量的方法进行反复调试，确定了《笨花》的叙事可提炼为两大核心且独立的主题，即主题数设置为 2（超过 2 生成的主题与其他主题重叠且权重极小）。迭代次数设置为 15，主题分布先验 alpha 设置为 0.6，并将结果可视化为下图 7-9。

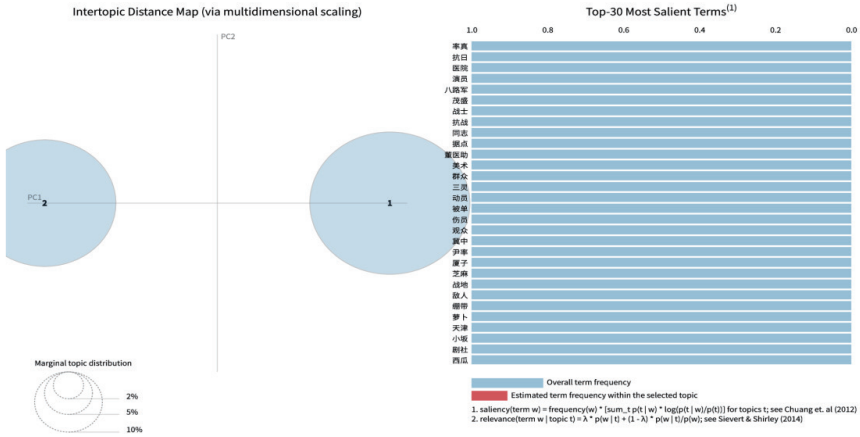

图 7

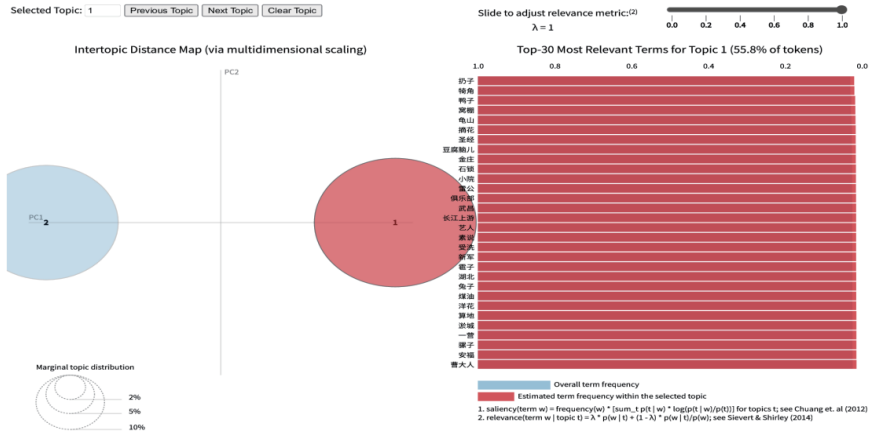

图 8

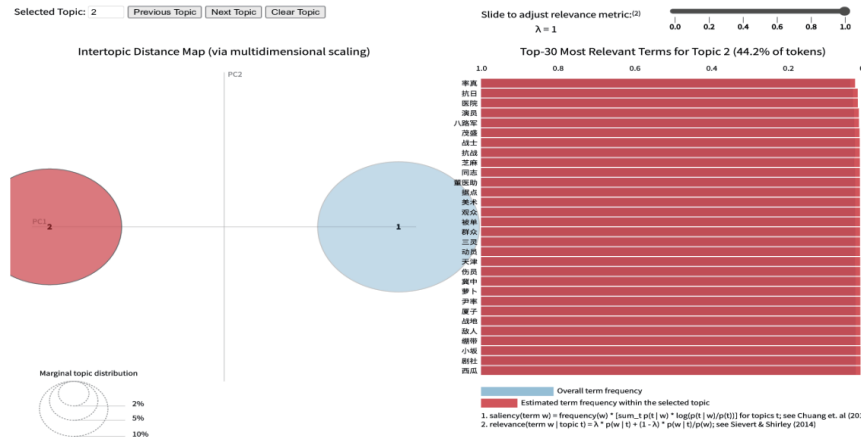

图 9

LDA 主题建模结果显示，《笨花》的叙事可提炼为“日常民生”（图 8）与“战争”（图 9）两大核心主题，二者区分度高、无明显重叠，且通过主题权重分布反映了不同叙事维度在文本中的优先级，即乡村生活略高于战争。两个主题的点在图中距离较远，是完全独立的两大叙事维度，说明在分析《笨花》的城市化进程这样一个贯穿全文的命题时，不能脱离战争背景单独讨论城乡互动，日常民生中的相当一部分元素，本质上是战争时期社会结构变化的产物。同时，日常民生叙事的占比更高，这说明《笨花》的叙事重心是日常民生，而战争是嵌套在此背景下的次级叙事，城市化进程的发展需结合战争加以理解，而不是孤立的社会经济现象。

为了反映不同阶段情感态度的变迁，首先用 Python 遍历各阶段文本，保留包含至少一个“城市化关键词”的句子，并保存为文件。接着加载

BosonNLP 情感词典与知网 Hownet 词典以作参考。随后，项目调用预训练模型 Erlangshen-Roberta-110M-Sentiment 对文本进行情感赋值。利用循环神经网络进行情感分析分别绘制了各阶段情感均值柱状图与各阶段情感分布饼图（图 10-11）：

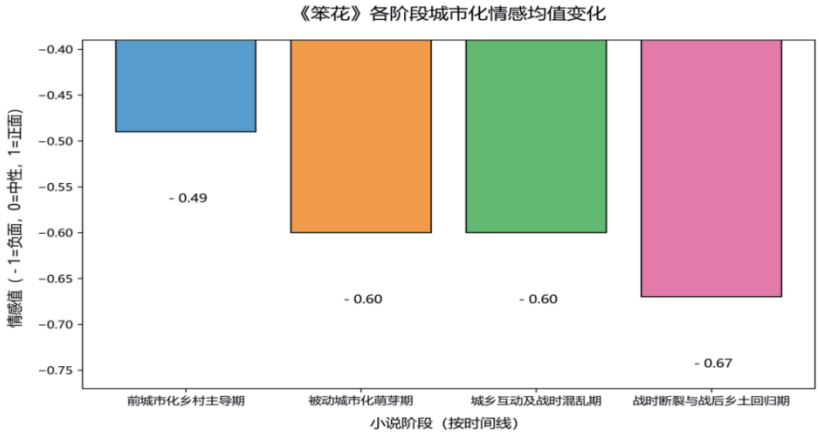

图 10　情感均值柱状图

柱状图覆盖四个阶段，所有阶段情感均值均为负值，说明《笨花》中与城市化相关的文本整体以负面情感为主，未出现正面主导的阶段。从时间线看，整体情感态度呈现负面程度逐步加深的线性趋势。

在前城市化乡村主导期（–0.49），负面程度最弱，接近中性偏负，此时城市化尚未明显渗透，乡村生活相对稳定，与城市化相关的冲突较少，负面情感仅为潜在状态。

在被动城市化萌芽期到城乡互动及战时混乱期（均为 –0.60），负面强度较阶段 1 显著上升并维持稳定——城市化开始渗透，伴随一系列负面情节，同时战时混乱加剧生存压力，负面情感成为主流。

在战时断裂与战后乡土回归期（–0.67），负面程度最强——战争导致城乡联系断裂，生存危机加剧，而战后乡土回归并未带来城市化红利，反而可能因城市化中断和战前、战后的生活落差强化负面认知，成为全阶段负面情感最集中的时期。

前城市化乡村主导期的情感分布饼图，核心呈现负面主导，正面仍存一定空间的结构特征。该阶段负面情感占比达 69.1%，虽此时城市化尚未对乡村形成明显渗透，文本以乡村日常生活描述为主，但负面情感已包含对未来城市化冲击的潜在隐忧；正面情感占比 20.0%，是四个阶段中正面占比最高的时期，这与阶段特征高度契合，城市化未实质介入，文本中大量描述乡村日常作息，人物与邻里关系的内容较为和谐正面，构成正面情感的主要来源；

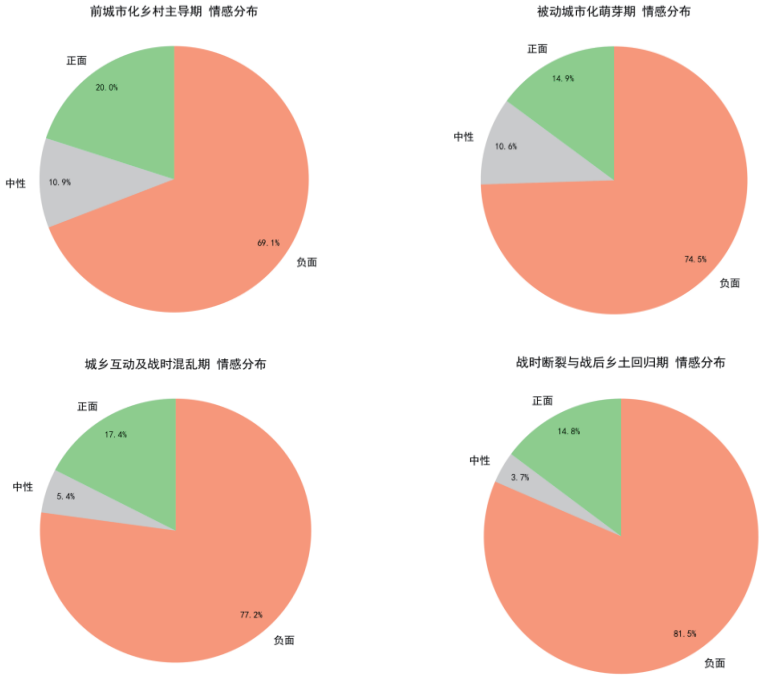

图 11　各阶段情感分布饼图

中性情感占比为 10.9%，与城市化无直接关联。

被动城市化萌芽期的情感分布饼图，呈现出负面占比上升，正面略微收缩的结构变化。随着城市化开始渗透，负面情感占比从 69.1% 升至 74.5%，继续成为阶段主导情感；中性情感占比几乎没有变化，正面情感占比下降至 14.9%，仍远低于负面。

城乡互动及战时混乱期的情感分布饼图，呈现负面占比再升、中性显著收缩的结构特点。此阶段城乡互动频次增加，随着战乱的影响，负面情节集中爆发，推动负面情感占比进一步攀升至 77.2%；中性情感占比仅 5.4%，几乎压缩至极致，因该阶段文本完全聚焦城市化相关的冲突与战时生存困境，无多余的客观中性描述，大部分内容均带有明确的情感倾向；正面情感占比升至 17.4%，相比上一阶段较高，但提升并不显著，无法改变负面主导的整体格局。

战时断裂与战后乡土回归期的情感分布饼图，呈现负面占比达峰值，中性趋近于零的结构特征。战时导致城乡联系断裂、城市化进程中断，人员伤亡、物资匮乏等问题加剧，战后乡土回归未带来预期的生活改善，使负面情感集中爆发，占比升至全阶段最高的 81.5%；中性情感占比仅 3.7%，为四个阶段最低，无论是对城乡现状的描述，还是对人物命运的记录，均带有强烈的负面情感色彩，客观中性的内容几乎消失；正面情感占比回落至 14.8%，相较于前一阶段的 17.4% 略有下降，正面情感持续时间短、覆盖范围窄。

负面情感的持续加剧，与《笨花》中城市化从潜在影响到直接冲击，再到战时中断的情节推进一致——城市化未带来预期红利。正面占比始终低迷（10.9%-17.4%），印证了《笨花》对城市化的批判性视角，小说并未大肆渲染城市化的积极面。

3.4 共现关系与文本相似度

本研究通过使用 Python 工具分别构建了四个阶段高频核心词的共现矩阵，并进行可视化分析（见图 12），词频越高，颜色越深。其目的在于通过共现揭示不同阶段的关联模式变化。

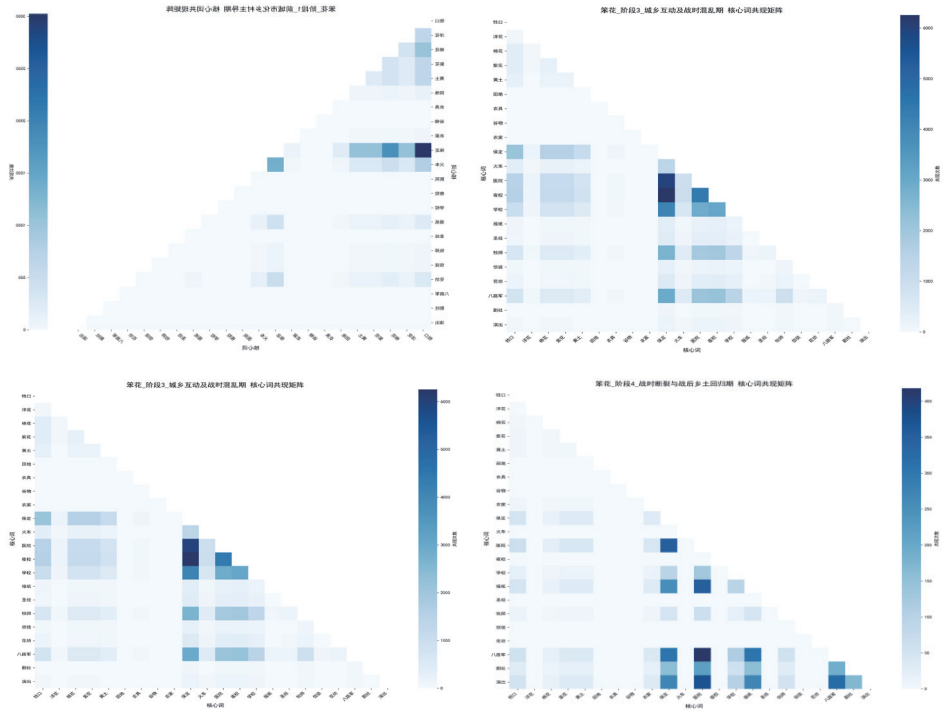

图 12　四阶段核心词共现矩阵图

从图示可见：第一阶段仅农村类核心词形成强共现（保定市作为城乡的接口），且集中在农业生产链条上。说明此阶段无城乡人员、资源流动，文本完全围绕乡村内部生产生活展开，符合前城市化的本质即乡村是独立且封闭的叙事单元，城市化要素尚未渗透。第二阶段中，医院、夜校、学校、火车等城市类核心词出现强共现，农村类核心词的共现强度较第一阶段有明显下降。这说明城市化虽然并非主动推进，但已经很大程度上动摇了乡村主导的格局。第三阶段从单一农村关联或城市关联转向农村类，城乡类，战时类的多元强共现，且是四阶段中共现强度最高、关联最复杂的阶段，说明战乱打破了城乡的封闭性，人口、资源、组织的跨城乡流动达到顶峰，因此核心

词关联最紧密。到了第四阶段，核心词的共现强度出现下降，战时类词如“八路军”并未失去关联。说明战乱即将结束，人口从混乱流动回归乡土，城乡互动因战时需求消失而减少，整体处于稳定但松散的状态，战争的影响依然巨大。四张热图的共现强度与关联主体变化，清晰反映了《笨花》中城市化进程并非线性的“农村—城乡互动—城市主导”，而是受战乱干扰的“孤立—萌芽—强互动”（战时—回归的波动格局）。

为了验证四阶段划分的可行性和关键词的匹配度，本研究首先将四个阶段的预处理文本（分词后的数据）导入 Python，并将其转为 TF-IDF 向量，接着计算余弦相似度并生成相似度矩阵（取值范围为 0-1，越接近 1 表示越相似），然后用 Seaborn 绘制热力图（见图 13），颜色越深表示相似度越高：

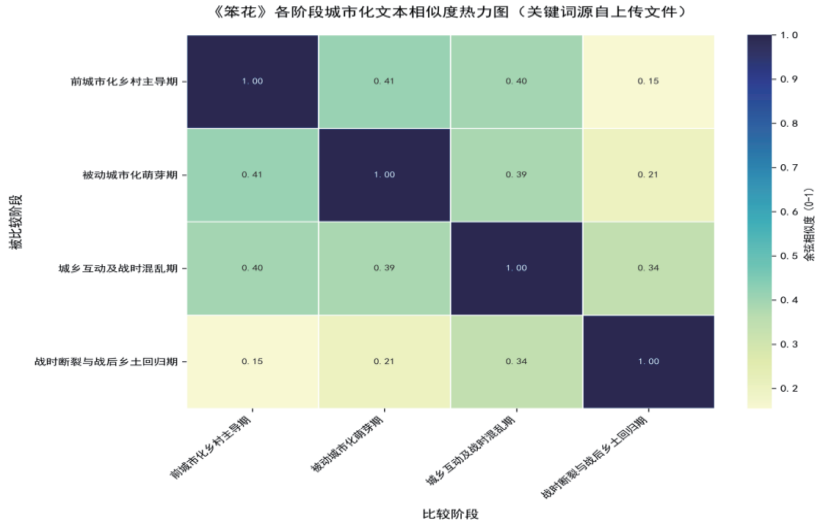

图 13　各阶段城市化文本相似度热力图

根据生成的热力图，可以看出，跨多阶段的间隔阶段相似度符合预期，这也验证了基础逻辑（及四个阶段的划分）的合理性。热力图中间隔 2 个及以上阶段的相似度普遍处于低位，说明四阶段文本的特征割裂明显，验证了阶段差异的真实性。同时，所有相邻阶段的相似度均低于 0.5，且呈逐阶段递减趋势，这反映了《笨花》的城市化进程并非平稳渐进，而是存在连贯性断裂。

阶段 1 与阶段 2 的相似度为 0.41，是相邻阶段中最高值，但仍低于 0.5。说明完全乡村到城市初步萌芽的过渡相对弱，城市化特征并不如此明显。阶段 2 与阶段 3 的相似度降至 0.39。这说明战时混乱的特殊背景打破了从萌芽到深化的正常逻辑，使城乡互动期的文本特征与被动萌芽期的初步城市化特征关联性减弱。阶段 3 与阶段 4 的相似度进一步降至 0.34，是相邻阶段中最低值。这说明战时断裂直接导致城市化进程倒退，而战后人口从城市回流乡村、重归

农耕，使战后回归期的文本特征呈现一种中间态，与城乡互动期的特征割裂，由于城市化与战争的影响，其特征又与阶段 1 大相径庭（仅 0. 15 的相似度），形成进程中断的文本信号。

文档相似度分析显示，所有相邻阶段相似度均低于 0. 5 且呈递减趋势，这表明《笨花》的城市化进程并非平稳渐进，而是存在连贯性断裂，每个阶段具有独特的文本特征。

四、讨论：从“远读”到“细读”的往复

本研究的计算分析结果，不仅宏观上勾勒出《笨花》叙事的总体趋势，更在关键数据转折点上引导我们回归文本内部，发现其微观的叙事肌理与情感结构。

4.1 关键词频的“断裂”：非线性的城市化叙事模式

本研究通过多维度的计算分析，揭示了《笨花》所呈现的城市化进程并非传统认知中的线性发展模式，而是呈现“孤立—萌芽—互动—回归”的波动格局。这一发现挑战了城市化即“农村单向变为城市”的简化理解，凸显了 20 世纪上半叶中国乡村城市化的特殊性与复杂性。

词频分析显示，“笨花”与“城市”的词频在三个阶段呈现出显著的消长关系，这直观地印证了小说城乡叙事重心的转移。然而，更值得注意的是第三阶段（1937-1945）的“断裂”。在此阶段，“城市”词频的增长趋势被强行中断，而“战争”“日本”“队伍”等词汇以前所未有的频率主导叙事。这种宏观数据上的“断裂”在文本微观层面表现为叙事逻辑的根本改变。例如，在第二阶段，一个人物进城可能伴随着对“洋楼”“电灯”“柏油马路”等现代性符号的细致描绘。但在第三阶段，同样的进城行为，其叙事焦点则完全转向了“城墙上的炮眼”“街头的日本兵”和“躲避轰炸的人群”。小说中向义等人物的个人命运与理想追求，被战争这一无法抗拒的宏大历史力量所裹挟、中断。计算结果在此并非简单呈现词汇变化，而是量化地证明了战争是如何作为一种压倒性的“叙事主体”，取代了“城市化”这一原本的内生发展逻辑。

文本相似度的阶段性断裂（相邻阶段相似度均低于 0. 5）与关键词演化的非连续性特征，共同指向一个重要结论：在战争背景下，城市化进程具显著的脆弱性和可逆性。第四阶段城市化元素的几乎完全消失，不仅是战时断裂的直接体现，更反映了传统乡土社会强大的回归力量。这种“断裂与回归”的叙事模式，为理解转型期中国社会的复杂性提供了文学维度的证据。

4.2 LDA 主题演变：从乡土伦理到战争政治

LDA 主题模型的结果进一步揭示了“断裂”的内在结构。模型识别出的“战争”与“日常民生”双主题结构，以及“八路军”等战时元素的高频出现，揭示了战争在城市化进程中的关键作用。战争不仅是外部干扰因素，更是重塑

城乡关系的结构性力量。

第一阶段的“乡土伦理”与“宗族关系”主题，在第二阶段逐渐让位于“城市发展”与“个人机遇”，这符合传统的现代化叙事。然而，第三阶段的主题模型再次出现剧变，“战争暴力”与“国家政治”成为压倒性主题。这种主题的强制性转向，在文本中体现为人物话语和思维方式的改变。例如，小说前半部分人物间的对话多围绕土地、收成、家族荣誉等展开，其行为逻辑根植于乡土伦理。进入战争时期，即便是最乡土的人物，其对话也开始充斥着“救国”“汉奸”“抗日”等政治话语。这种转变并非人物内在思想的有机成长，而是一种外部话语的强行植入。LDA 模型捕捉到的，正是这种乡土社会内在价值体系在国家主义宏大叙事冲击下的瓦解与重构。

从共现矩阵的变化可以看出，战时混乱期（第三阶段）呈现最复杂的多元共现模式，战争打破了城乡的封闭性，促进了人口、资源、组织的跨城乡流动。然而，这种战时驱动的城乡互动具有临时性和非持续性特征，一旦战争结束，城市化进程即出现明显退潮。这一发现对理解中国现代化进程中的断裂与延续具有重要启示意义。

4.3 情感分析的复杂性：在负面基调中挖掘“反常”

循环神经网络（RNN）的情感分析揭示的持续负面倾向（−0.49 至 −0.67 的递进式加深），说明小说整体呈现出显著的负面情感基调，并在第三阶段（战争时期）达到顶峰。这印证了小说对城市化进程的批判性反思态度，以及战争带来的巨大创伤。这种批判并非简单的反城市化，而是对城市化过程中乡土价值流失、传统秩序瓦解的深刻忧虑。正面情感占比的持续低迷（始终低于 20%）表明，作品并未渲染城市化带来的现代性红利，而是着重呈现转型过程中的痛苦与迷茫。中性情感在战后回归期降至 3.7% 的极低水平，说明在经历战争与城市化的双重冲击后，文本叙事完全被强烈的情感色彩所主导，客观中立的描述几乎消失。

然而，一个值得深究的“反常数据”是：尽管第三阶段的负面情感值最高，但正面情感值相较于第二阶段末期，出现了一个微弱但清晰的回升。传统的阐释可能会将此归为统计误差或忽略不计。但从人文研究的深度来看，这恰恰可能是文本复杂性的关键所在。这个微弱的正面情感回升，揭示了一种矛盾心态：战争的暴力摧毁了城市现代化的虚假繁荣，同时也冲击了压抑人性的传统宗法制度。对于小说中的某些边缘人物而言，这种混乱可能短暂地提供了一种摆脱旧有束缚的“机会”或“希望”，尽管这种希望本身是脆弱和虚幻的。

同时，面对外敌，“抗日”“救亡”等集体主义话语在一定程度上取代了第二阶段中个人主义的疏离与焦虑。文本中对“万众一心”或“一致对外”的短暂、理想化描绘，会瞬间激发正面情感，即便其背景是惨烈的战争。从这一“反常数据”，我们看到《笨花》的情感结构并非单向度的悲观主义，而

是呈现出内在张力与矛盾的复杂性。计算方法帮助我们精确定位了这一情感的“微澜”，而深入的文本细读则揭示了这“微澜”背后复杂的历史与人性内涵。这种情感取向与铁凝一贯的人文关怀相契合，体现了作家对底层民众命运的深切关注。

4. 4 方法论的贡献与局限

本研究展示了文学计算批评在分析复杂文学现象中的独特价值。通过词频统计、主题建模、情感分析等计算方法的综合运用，我们得以从宏观视角把握《笨花》城市化叙事的整体特征，发现了传统细读方法难以察觉的潜在模式。尤其是文本相似度分析和对应分析的运用，为验证文学直觉提供了量化证据。

然而，本研究也存在一定局限。首先，阶段划分虽基于叙事逻辑，但仍具有一定主观性。其次，自定义词典的构建可能存在遗漏，影响分析的完整性。第三，情感分析模型的准确性在处理文学文本的复杂修辞时可能存在偏差。未来研究可以通过扩大样本范围、优化算法模型、结合更多维度的分析方法来进一步深化认识。

五、结语

本研究证实，综合运用词频统计、LDA 主题模型和 RNN 情感分析等计算方法，有效揭示长篇小说《笨花》中隐含的城市化进程的复杂叙事模式与情感态度。在实证层面，量化并呈现了《笨花》中一个“断裂与回归”的非线性城市化叙事，并指出了“战争”是导致这一断裂的核心外部变量。在理论层面，将《笨花》的文本发现与“被中断的现代性”理论和叙事学理论进行链接，论证了宏大历史对文学微观叙事的重构机制，并在中国乡土文学脉络中确立了其独特的批判性地位。

《笨花》所呈现的城市化并非通往繁荣的康庄大道，而是一个充满创伤、被外部力量（战争）强行打断的现代化进程。小说最终“回归”乡土，并非田园牧歌式的浪漫想象，而更像是一种在城市梦想破灭后的无奈选择与文化自救。本研究的数据——特别是城市相关词频的增长停滞和负面情感的显著上升——为这一理论判断提供了量化支持，揭示了中国特定历史情境下现代性经验的复杂性与断裂性。

传统叙事学分析历史与叙事的关系，多依赖于定性的文本解读。而本研究则从词汇、主题和情感三个维度，量化了历史事件对叙事“微观结构”的塑造力。战争不仅改变了故事的“情节”，更深刻地改变了“话语”——它改变了人物的口头语言（LDA 主题），重塑了文本的情感基调（RNN 情感值），甚至改变了叙事聚焦的空间符号（词频）。这为理解“历史如何进入文本”这一核心叙事学命题，提供了一个具体、可操作的计算分析范例。

如果将《笨花》置于更广阔的中国当代文学史中，我们可以发现：与路遥

《平凡的世界》中那种虽然充满苦难但仍对城市充满向往、最终实现个人价值的“奋斗叙事”不同，《笨花》提供了一个截然相反的样本。本研究揭示的压倒性负面情感，表明在对城市化进程的态度上，《笨花》有别于主流的“改革与发展”叙事，而更接近于一种对现代性代价进行深刻反思的批判性传统。它所展示的“断裂”，并非个人奋斗的挫折，而是整个民族历史的创伤印记。

最终，本研究倡导一种“往复式”的数字人文研究路径：始于人文问题的提出，借由“远读”的计算方法发现宏观模式与“反常”数据点，再“往复”到文本内部进行“细读”阐释，最终将“远读”与“细读”的发现共同整合到与核心理论的对话之中。这不仅是从“我用计算机发现了什么”到“我的发现对人文知识意味着什么”的转变，更是确保数字人文研究能够真正产生深刻洞见、贡献理论新知的核心所在。这种循环往复的过程，确保了数字人文研究不会沦为冰冷的技术展示，而是始终由人文问题驱动，并最终回归到人文知识的创造与理论创新之中。它既是本研究的实践路径，也为我们未来从事相关计算文学研究提供了一个可资借鉴的范型。这种方法论的自觉，是推动数字人文从技术辅助走向理论创新的关键一步。

Works Cited

Blei, David M., Andrew Y. Ng and Michael I. Jordan. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3 (2003): 993-1022.

BosonNLP 情感词典 . Available at: https://bosonnlp.com/. Accessed 21 Sept. 2025.

“Erlangshen-RoBERTa-110M-Sentiment.” Hugging Face. Available at: https://huggingface.co/IDEA-CCNL/Erlangshen-RoBERTa-110M-Sentiment. Accessed 21 Sept. 2025.

Gensim: Topic Modelling for Humans. Available at: https://radimrehurek.com/gensim/. Accessed 21 Sept. 2025.

HowNet 知网义原词典 . Available at: https://www.keenage.com/. Accessed 21 Sept. 2025.

“Jieba 中文分词 .” GitHub. Available at: https://github.com/fxsjy/jieba. Accessed 21 Sept. 2025.

Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. Urbana-Champaign: U of Illinois P, 2013.

Moretti, Franco. *Distant Reading*. London, New York: Verso, 2013.

Prince. Python Factor Analysis Library (CA, MCA, FAMD). Available at: https://github.com/MaxHalford/prince. Accessed 21 Sept. 2025.

铁凝：《笨花》。长沙：湖南人民出版社，2008 年。

[Tie Ning. *Clumsy Flower*. Changsha: Hunan People’s Publishing House, 2008.]

Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: U of Chicago P, 2019.

Voyant Tools. Available at: https://voyant-tools.org/. Accessed 21 Sept. 2025.