

《狼王梦》词频统计和词汇分析

姓名：董欣然

学号：1900013018

(一) 简介

《狼王梦》全文约 151000 字，是一部以自然主义观点叙事的动物小说。讲述了母狼紫岚如何培育三小只公狼成为狼王的经过，把在大自然挣扎求生存的狼，及因求生存而发展出来的“狼道”，刻画得淋漓尽致。

选择《狼王梦》的原因是：我从小十分喜爱动物小说，尤其喜欢狼性。为了分析小说中的动物角色的重要程度和儿童文学中的遣词套路，我将角色和非角色的词汇分为两类，分别对词频展开分析。

(二) 数据处理

- 发现标点符号、空格、“的”、“了”等单字出现频率极高，为了排除这些无实意的单字，作业中去掉了所有单字。
- 发现小说中的重要角色不在 jieba 的自带词库中，例如“洛戛”“大白狗”，故将其添加进自定义词库。
- 发现主人公“紫岚”，有时会被划分成词汇“当紫岚”、“紫岚想”，故将“紫岚”添加进自定义词库。这样保证了一定能划分出单词“紫岚”。
- 同理将主人公“媚媚”添加进自定义词库。这样保证了一定能划分出单词“媚媚”。
- 通过整理词汇表，可知小说中的角色十分有限，故将所有角色记录在 character_list 中，方便与非角色词汇区分。

(三) 代码功能

- 将角色和非角色的词汇分为两类，可选择展现其中任意一类。
- 将词汇按词频从多到少排序，可根据需求输出排名最高的若干个词汇。
- 将查询结果导入到 excel 文档中。

(四) 结果分析

通过对角色词频的分析可知：

- 词频最高的是“紫岚”，“紫岚”是小说的主人公，全文是以“紫岚”的视角叙事的，词频最高非常合理。
- 按词频顺序接下来依次是“媚媚”，“双毛”，“蓝魂儿，这几个角色都是“紫岚”的子女，“紫岚”将培育狼王的梦想寄托在她的子女们身上，所以这些角色也是重中之重。
- 还有词频较高的是“老雕”。“老雕”出现在篇末，小说用巨大的篇幅描述了“紫岚”和“老雕”殊死搏斗的场景，最后紫岚死在老雕的爪下，全文告终。

- 通过分析可知，出现频次较高的角色和故事情节的发展息息相关，正是这些主要角色组成了全文的脉络。

通过对非角色词频的分析可知：

- “自己”出现的频次最高，“一种”、“一个”、“一声”、“一只”、“一匹”等量词的词频较高，说明这篇小说用词简单易懂，方便儿童阅读。
- “狼群”、“母狼”、“公狼”的词频较高，说明这是一篇与狼息息相关的小说，与事实相符。