

基于语料库的《哈利·波特与魔法石》与 《哈利·波特与死亡圣器》文本对比分析

陈 柱, 赵 冰

江西理工大学外国语学院, 江西 赣州

收稿日期: 2025年6月9日; 录用日期: 2025年7月14日; 发布日期: 2025年7月25日

摘 要

本文基于语料库的方法, 对J.K. Rowling创作的《哈利·波特》系列小说中的首部作品《哈利·波特与魔法石》和最终部《哈利·波特与死亡圣器》进行了对比分析。研究构建了一个总词数超过156万的英语可比语料库, 从词汇、句法和语篇等层面对两部文本进行量化分析。借助Wordsmith Tools 7.0、AntConc 3.5.9以及Readability Analyzer等工具, 完成了词性标注、词频统计、句式结构测量及可读性评估。研究结果显示, 随着叙事的发展和读者群体的成长, 作者在语言风格上呈现出明显变化, 包括词汇复杂度提高、平均句长增加及衔接手段使用的差异。这些变化不仅体现了作者语言风格的演进, 也为文学文体学研究和基于语料库的叙事分析提供了有价值的参考。

关键词

语料库, 文体分析, J.K. Rowling, 《哈利·波特》系列

A Corpus-Based Comparative Analysis of *Harry Potter and the Philosopher's Stone* and *Harry Potter and the Deathly Hallows*

Zhu Chen, Bing Zhao

School of Foreign Languages, Jiangxi University of Science and Technology, Ganzhou Jiangxi

Received: Jun. 9th, 2025; accepted: Jul. 14th, 2025; published: Jul. 25th, 2025

Abstract

This study conducts a corpus-based comparative analysis of *Harry Potter and the Philosopher's*

文章引用: 陈柱, 赵冰. 基于语料库的《哈利·波特与魔法石》与《哈利·波特与死亡圣器》文本对比分析[J]. 现代语言学, 2025, 13(7): 861-872. DOI: 10.12677/ml.2025.137781

Stone and Harry Potter and the Deathly Hallows, the first and final books in J.K. Rowling's *Harry Potter* series. By constructing a comparable English corpus comprising over 1.56 million words, the research quantitatively examines differences in lexical, syntactic, and discourse features across the two texts. Tools such as Wordsmith Tools 7.0, AntConc 3.5.9, and Readability Analyzer were used for text annotation, frequency analysis, sentence structure measurement, and readability evaluation. The findings reveal significant shifts in linguistic style, including increased lexical complexity, longer average sentence lengths, and variations in the use of cohesive devices, reflecting the narrative progression and maturation of both the author and her audience. This study highlights the evolving stylistic characteristics in Rowling's writing and offers insights into literary stylistics and corpus-based narrative analysis.

Keywords

Corpus, Stylistic Analysis, J.K. Rowling, *Harry Potter* Series

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. 研究问题

本研究主要基于 J. K. Rowling 创作的《哈利·波特》系列小说的第一部《哈利·波特与魔法石》(*Harry Potter and the Sorcerer's Stone*)和最后一部,即第七部《哈利·波特与死亡圣器》(*Harry Potter and the Deathly Hallows*)的语言特征进行比较,前者于 1997 年 6 月 26 日在英国出版,后者出版于 2007 年。本文主要从词汇、句法和语篇三个层面,对《哈利·波特与魔法石》和《哈利·波特与死亡圣器》的英语文本进行文体特征的比较与分析。量化分析随着时间的推移,作者的语言使用是否发生了变化,以及这些变化是如何影响整体故事叙述的词汇与句法特征。

2. 研究过程

本研究旨在通过构建一个涵盖《哈利·波特》系列小说中首部作品《哈利·波特与魔法石》(*Harry Potter and the Philosopher's Stone*)与最后一部作品《哈利·波特与死亡圣器》(*Harry Potter and the Deathly Hallows*)的英语可比语料库,对两部作品的文体特征进行定量与定性的综合分析。为确保数据的可靠性与处理过程的科学性,研究严格遵循以下步骤:

研究首先从 Z-Library 电子书平台获取上述两部作品的英文原版 PDF 格式电子书。为便于后续文本处理与格式转换,采用 Solid Converter PDF 软件将 PDF 文件转换为可编辑的 Word 文档。该步骤确保了文本的完整性与连续性,并为进一步的数据清洗和格式标准化打下基础。

将获得的 Word 文档导入 Sublime Text 编辑器中,并统一转换为 TXT 纯文本格式。该格式在语料库分析软件中具有较高的兼容性和处理效率,有利于后续的词频提取、关键词分析和句法结构研究。同时,为避免编码错误与符号缺失,研究者对文本编码格式进行了统一设置,确保全部文件均采用 UTF-8 标准。

在文本转换完成后,研究者对语料文本进行了系统的清洗和预处理。主要包括以下几个方面:(1) 去除原始文本中的页眉、页脚、章节编号等非正文内容;(2) 删除图片说明、插图描述、图表注释等非语言性信息;(3) 清除脚注、尾注等可能影响语言分析的数据干扰项。此过程以确保保留的文本数据仅为小说正文部分,从而提高语料分析的准确性与代表性。

为进一步提取语言特征并进行文体分析,研究借助 Free CLAWS Web Tagger 在线标注工具对清洗后

的文本进行了 C7 词性标注。该工具基于英国拉夫堡大学开发的 CLAWS 系统, 能够为文本中的每一个词语赋予相应的词性标记(如名词 NN、动词 VV、形容词 JJ、副词 RR 等), 为后续的词性分布统计、句法结构分析及风格特征比对提供了坚实的数据基础。

在完成上述处理步骤之后, 正式建立了一个基于《哈利·波特》系列首尾两部小说的英语可比语料库。该语料库共包含两篇完整小说文本, 总词汇量达 1,562,237 词, 具有篇幅长、体裁统一、风格连贯的特点。该语料库作为研究的核心数据源, 在保证真实语料代表性的基础上, 便于实现纵向比较和横向分析。

本研究综合使用了以下三款语料库工具: Wordsmith Tools 7.0: 用于提取关键词、分析词频分布、构建词汇列表等; AntConc 3.5.9: 用于分析词汇搭配、语境共现、关键词语料行(KWIC)等; Readability Analyzer: 用于评估文本的可读性水平, 辅助文体特征分析。

以上软件各具功能, 协同使用能够从多个维度对文本进行量化分析与文体特征探测, 为本研究所采用的定量语料库方法提供了有力的技术支持。

3. 研究结果

3.1. 词汇特征对比分析

对于词汇特征的研究, 从平均词长、词汇密度、类符与形符比率、词目类型、高频词频率参数这五个方面进行总体描述。

3.1.1. 词汇长度

平均词长指文本中所使用词语的平均长度, 以词语的字母数为计算标准。词长标准差指文本中每个单词长度与平均词长的差异[1]。一般而言, 平均词长的数值越高, 表明文本中使用的复杂词就越多。将《哈利·波特与魔法石》与《哈利·波特与死亡圣器》的文本分别导入 Wordsmith7.0, 运行 Word List 功能, 得到两个语料子库的平均词长和词长标准差。从表 1 数据可得, 《哈利·波特与死亡圣器》的平均词长和词长标准差均高于《哈利·波特与魔法石》, 表明前者的用词复杂程度和正式程度相对较高。

Table 1. Average word length and word length standard deviation of the two corpora

表 1. 两个语料子库的平均词长及词长标准差

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
平均词长	4.34	4.48
词长标准差	2.05	2.32

另外, 英语中通常将 6 个字母以上的词称为长词。表 2、表 3 数据表明, 《哈利·波特与死亡圣器》的长词占比高于《哈利·波特与魔法石》, 说明《哈利·波特与死亡圣器》文本使用的长词较多, 在词汇复杂度上相较于《哈利·波特与魔法石》有所提升。这种提升不仅体现在长词的使用频率上, 还反映两部作品中故事情节、主题深度以及读者受众的变化。在《哈利·波特与魔法石》中, 作为系列的第一部作品, 作者 J. K. Rowling 更多地采用了简单明了的词汇来构建哈利·波特的魔法世界, 以便于吸引不同年龄层的读者群体, 特别是年轻读者。而到了《哈利·波特与死亡圣器》, 作为系列的终结篇, 故事情节变得更加复杂, 涉及了更多深层次的情感纠葛和道德选择, 这就需要更多的长词来准确表达这些复杂的概念和情感。此外, 长词的使用也与作者的写作风格有关。随着系列的进展, J. K. Rowling 的写作技巧更加趋于成熟, 也更倾向于使用长词来丰富文本的语言表达, 增强故事的表现力。

Table 2. Total number and proportion of long words in the two corpora
表 2. 两个语料子库的长词总数与占比

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
长词	12,190	39,350
总词数	77,916	205,920
长词占比	15.65%	19.11%

Table 3. Vocabulary statistics table
表 3. 词汇统计表

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
1-letter words	4941	2449
2-letter words	37,093	11,614
3-letter words	43,831	17,800
4-letter words	36,732	15,564
5-letter words	25,493	10,415
6-letter words	18,480	7884
7-letter words	15,109	5569
8-letter words	10,414	3452
9-letter words	6453	1808
10-letter words	3920	977
11-letter words	1623	234
12-letter words	1072	113
13-letter words	416	26
14-letter words	234	5
15-letter words	62	5
16-letter words	34	0
17-letter words	10	0
18-letter words	3	1

3.1.2. 词汇密度

词汇密度指文本中的词项数量(即实词数量)与该文本的单词总量之比,计算方法为:词汇密度 = 实词数/词汇总数 $\times 100\%$ 。英语中的实词指名词、实义动词、形容词和副词[2]。利用 Free CLAWS web tagger 网页对《哈利·波特与魔法石》和《哈利·波特与死亡圣器》两个语料子库进行词性标注后,运行 AntConc3.5.9 的 concordance 功能分别利用正则表达式检索所有的名词:\S+_N\w+、所有的动词:\S+_V\w+、形容词:\w+_JJ 和副词:\w+_RR, 可得两个语料子库的名词、实义动词、形容词、副词以及实词总数,随后根据计算公式得出词汇密度。

Table 4. Total number of content words and lexical density of the two corpora
表 4. 两个语料子库的实词总数及词汇密度

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
名词	18,032	22,505
实义动词	19,051	22,534
形容词	4009	5631
副词	3618	4415
实词总数	44,710	55,085
总词数	77,916	205,920
词汇密度	57.38%	26.75%

根据表 4，《哈利·波特与魔法石》词汇密度数值高于《哈利·波特与死亡圣器》，表明其实词占比较多。较高的词汇密度意味着作者使在《哈利·波特与魔法石》中使用了更多的实词来描述魔法世界的细节、人物的性格以及故事情节的发展。这种词汇使用方式有助于读者更具体地理解故事背景和角色行为，同时也增强了文本的表现力和可读性。相比之下，《哈利·波特与死亡圣器》作为系列的最后一部作品，更加注重情感和情节的深度挖掘，因此在词汇使用上更加倾向于使用抽象词汇和隐喻等修辞手法。这种写作风格使得文本在表达上更加含蓄和深刻，但同时也导致了词汇密度的降低。

3.1.3. 类符/形符比

语料库语言学中，类符是文本中所使用的不同词汇的种类，形符是文本中所有词汇的总数量。类符/形符比(type/token ratio, TTR)是指文本中所使用的不同词语的数量与词语总数量间的比值，计算公式为类符/形符比 = (类符数/形符数) × 100%。由于常用的不同词汇数量有限，文本长度会有较大的差异，因此 Scott 提出采用标准化类符/形符比(standardized type/token ratio, STTR)作为计量标准，计算方式以一个文本中按每 1000 词算出一个 TTR，然后再取所有 TTR 的平均值。[1]标准化类符/形符比值越大，表示该文本词汇重复率低，使用不同词汇的数量越多，词汇变化性和多样性程度高，反之亦然。根据 Wordsmith7.0 中的 Word List 功能，研究者得到两个语料子库的类符与形符数及标准类符与形符比率。

Table 5. Type-token ratio of the two corpora
表 5. 两个语料子库的类符与形符比

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
类符	5966	12,125
形符	77,916	205,920
标准化类/形符比(%)	43.93%	44.54%

表 5 数据表明，《哈利·波特与魔法石》文本的标准化类符/形符比低于《哈利·波特与死亡圣器》，说明《哈利·波特与魔法石》所使用的不同词汇量相对较小，词汇重复率高，词汇多样性程度低于《哈利·波特与死亡圣器》。这一发现反映了作者 J.K. Rowling 在创作两部作品时的不同策略和风格。在《哈利·波特与魔法石》中，作为系列小说的开篇之作，作者更侧重于构建魔法世界的基础设定，介绍主要角色和情节，因此在词汇使用上相对保守，采用了更多重复和常见的词汇，以便于读者理解和接受。然而，在《哈利·波特与死亡圣器》中，作为系列小说的终结篇，作者需要处理更加复杂和丰富的情节，同时还需要对整个系列进行回顾和总结。因此，作者采用了更多不同的词汇来表达复杂的思想和情感，增

加了文本的词汇多样性。

3.1.4. 词目类型

词目类型是指一个词汇的基本形式(lemma)在语言中存在的不同词形或变体。这些变体通常包括时态、语态、单复数等形态上的变化[3]。Lemma-Word Ratio (词目 - 词汇比率)是指在一个文本或语料库中, 词目(lemma)的数量与总词汇(word)数量之间的比率。Lemma-Word Ratio 反映了文本或语料库中词汇的丰富程度和多样性。比率越高, 说明文本或语料库中的词汇变化越多, 涉及更多的语法结构和语义内容。在语言学研究、自然语言处理、文本分类等领域, Lemma-Word Ratio 可以作为衡量文本复杂性和多样性的一个重要指标。Lemma TTR 是指在一个文本或语料库中, 不同词目(lemma)的类型数(Types)与词目总数(Tokens)的比率。Lemma STTR 是对 Lemma TTR 进行标准化处理后的指标, 旨在解决文本长度对 TTR 值的影响。由于 TTR 值会受到文本长度的影响(通常文本越长, TTR 值越低), 因此通过标准化处理可以使不同长度的文本在词汇丰富度上具有可比性。

将《哈利·波特与魔法石》与《哈利·波特与死亡圣器》的文本分别导入 Readability Analyzer 运行 Analyze 功能, 得到如下分析结果:

Table 6. Word lists of the two corpora
表 6. 两个语料子库的词目

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
Lemma Types	4248	8482
Lemma-Word Ratio	0.7191	0.7227
Lemma TTR	0.0529	0.0432
Lemma STTR	0.0624	0.0528

根据表 6 数据可得:

《哈利·波特与魔法石》的词目类型数为 4248, 而《哈利·波特与死亡圣器》的词目类型数为 8482。这显示了《死亡圣器》使用了更多不同的基本词汇形式(即词目)。这是因为《死亡圣器》作为系列的最后一部, 包含了更复杂的情节、更多的角色和更深入的背景设定, 因此需要更多的词汇来表达这些复杂的概念。

《哈利·波特与魔法石》的词目 - 词汇比率为 0.7191, 而《哈利·波特与死亡圣器》的比率为 0.7227。这意味着两部作品中, 大约 72% 的词汇都是不同的词目。这一比率在两部作品中非常接近, 表明两部作品在词汇使用的丰富性上大致相当。尽管《死亡圣器》的词目类型数更多, 但其总词汇量也相应增加, 因此词目 - 词汇比率并未显著变化。

《哈利·波特与魔法石》的词目 TTR 为 0.0529, 而《哈利·波特与死亡圣器》的 TTR 为 0.0432。这一比率通常用于衡量文本的词汇多样性。较低的 TTR 值表明文本中重复使用的词目较多, 而较高的 TTR 值则表明文本使用了更多的不同词目。在这里, 《魔法石》的 TTR 值略高, 因为该书的篇幅相对较短, 且主要介绍了哈利·波特世界的基本设定和角色, 因此需要使用更多的不同词目来传达这些新信息。而《死亡圣器》由于情节更加复杂, 需要更多地重复使用一些关键词目来强化主题和角色。

《哈利·波特与魔法石》的 STTR 为 0.0624, 而《哈利·波特与死亡圣器》的 STTR 为 0.0528。与 TTR 相比, STTR 考虑了文本长度的影响, 使得不同长度的文本在词汇多样性上具有可比性。在这里, 《魔法石》的 STTR 值仍然略高于《死亡圣器》, 进一步支持了上述关于 TTR 的分析。这表明《魔法石》在较短篇幅内使用了相对较多的不同词目来介绍新信息和构建世界设定。

综上所述, 这些参数反映了《哈利·波特与魔法石》和《哈利·波特与死亡圣器》在词汇使用上的差异和复杂度。尽管两部作品在词目 - 词汇比率上相近, 但《死亡圣器》使用了更多的不同词目类型, 这是由于其更复杂的情节和背景设定所需。同时, 《魔法石》在较短篇幅内使用了相对较多的不同词目来构建世界设定和介绍新信息。

3.1.5. 高频词参数

高频词指反复出现的一定数目的相同词汇, 即语料库中出现频率较高的词。高频词定义为“一个词项出现频率至少占库容的 0.10% 以上的词” [4]。参照此定义, 设定高频词所占比例的数值 $\geq 0.10\%$ 。借助 Wordsmith7.0 制作词表, 查询 Frequency 列表, 得出《哈利·波特与魔法石》和《哈利·波特与死亡圣器》的高频词数据。累计比例即高频词占整个词汇量的比例, 需要计算高频词在文本中出现的总次数, 将高频词的总次数除以整个文本的总词汇量, 由此得到累计比例。

Table 7. High-frequency word data of the two corpora

表 7. 两个语料子库的高频词数据

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
高频词数目	163	133
累计比例	58.76%	55.92%
高频词与低频词之比	0.0281	0.0110

Table 8. High-frequency words of the two corpora

表 8. 两个语料子库的高频词

哈利·波特与魔法石				哈利·波特与死亡圣器			
N	Word	Freq.	%	N	Word	Freq.	%
1	THE	3630	4.66	1	THE	10259	4.98
2	AND	1924	2.47	2	NOT	8564	4.16
3	TO	1861	2.39	3	AND	5330	2.59
4	A	1691	2.17	4	TO	4895	2.38
5	HE	1528	1.96	5	OF	4140	2.01
6	OF	1267	1.63	6	HE	3909	1.90
7	HARRY	1214	1.56	7	A	3538	1.72
8	WAS	1186	1.52	8	WAS	2749	1.33
9	IT	1027	1.32	9	HARRY	2717	1.32
10	IN	968	1.24	10	HIS	2615	1.27
11	HIS	937	1.20	11	IT	2545	1.24
12	YOU	885	1.14	12	IN	2230	1.08
13	SAID	794	1.02	13	HAD	1998	0.97
14	HAD	702	0.90	14	THAT	1937	0.94
15	I	654	0.84	15	SAID	1932	0.94
16	ON	637	0.82	16	YOU	1904	0.92

续表

17	AT	625	0.80	17	AS	1411	0.69
18	THAT	603	0.77	18	AT	1405	0.68
19	THEY	597	0.77	19	I	1304	0.63
20	AS	526	0.68	20	HIM	1301	0.63
21	HIM	501	0.64	21	NOT	1148	0.56
22	BUT	485	0.62	22	WITH	1131	0.55
23	WITH	416	0.53	23	ON	1115	0.54
24	RON	410	0.53	24	THEY	1109	0.54
25	ALL	398	0.51	25	HERMIONE	1074	0.52
26	OUT	375	0.48	26	HER	1044	0.51
27	UP	371	0.48	27	RON	1033	0.50
28	FOR	370	0.47	28	FOR	1004	0.49
29	BE	368	0.47	29	BUT	979	0.48
30	WHAT	344	0.44	30	SHE	929	0.45
31	HAGRID	336	0.43	31	FROM	841	0.41
32	THEM	325	0.42	32	THEM	813	0.39
33	WERE	306	0.39	33	HAVE	798	0.39
34	HAVE	297	0.38	34	BE	765	0.37
35	THERE	269	0.35	35	WERE	758	0.37
36	BACK	261	0.33	36	OUT	739	0.36
37	HERMIONE	257	0.33	37	UP	671	0.33
38	ONE	256	0.33	38	ALL	669	0.32
39	THIS	249	0.32	39	INTO	657	0.32
40	IF	245	0.31	40	THERE	653	0.32
41	FROM	241	0.31	41	COULD	642	0.31
42	SO	238	0.31	42	WHAT	604	0.29
43	NOT	237	0.30	43	BEEN	585	0.28
44	SHE	229	0.29	44	WAND	582	0.28
45	ABOUT	225	0.29	45	WHO	554	0.27
46	AN	219	0.28	46	IS	549	0.27
47	INTO	219	0.28	47	BACK	540	0.26
48	ME	219	0.28	48	KNOW	540	0.26
49	THEIR	217	0.28	49	NOT	536	0.26
50	KNOW	212	0.27	50	WE	515	0.25

根据表 7、表 8, 两个语料子库高频词的数目存在一定差距, 《哈利·波特与魔法石》的高频词累计比例要高于《哈利·波特与死亡圣器》子库(58.76% > 55.92%), 在《哈利·波特与魔法石》中, 高频词的

使用更为集中, 占据了更大的文本比例。作为系列小说的开篇之作, 《哈利·波特与魔法石》需要频繁使用一些核心词汇来构建世界观、介绍角色和设定情节, 因此高频词的使用更加频繁。同时《哈利·波特与魔法石》子库的高频词与低频词之比也要高于《哈利·波特与死亡圣器》子库($0.0281 > 0.0110$), 反映了作者在两部作品中词汇分布策略的不同。

3.2. 句式特征对比分析

针对句式特征的研究, 则从平均句长复合句特征、句子结构类型以及文本可读性等方面进行分析。

3.2.1. 平均句长

平均句长指一个篇章中的句子含有词语数量的平均值, 计算公式为: 平均句长 = 形符数/句子数[5]。句长标准差指句子的长度在平均句长左右浮动的程度, 标准差值越高, 表明文本中句子长短变化越大, 句式更为灵活, 可读性也就越强。平均句长和句长标准差是常用来衡量作者文体风格的参数类型。句子长度与可读性存在相关性。借助 Wordsmith7.0 的 word list 功能中的 statistics, 得出两个语料子库的平均句长。

Table 9. Average sentence length data of the two corpora

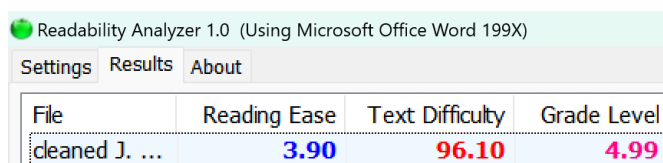
表 9. 两个语料子库的平均句长数据

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
句子个数	6311	6628
平均句长	12.34	31.06
句长标准差	9.92	28.52

表 9 数据显示, 《哈利·波特与死亡圣器》的平均句长数值大于《哈利·波特与魔法石》的平均句长数值, 说明《哈利·波特与死亡圣器》文本使用长句更为频繁。此外, 《哈利·波特与魔法石》的句长标准差数值低于《哈利·波特与死亡圣器》的句长标准差, 说明《哈利·波特与魔法石》的句子长短变化幅度小, 句式不如《哈利·波特与死亡圣器》灵活多变, 可读性相对较差。李长栓指出“简明英语”的两项基本原则: 使用简单句型以及限制句子长度。结合“第六声”更常用长句以及句子长短变化程度低的特点, “第六声”需多考虑句子内部的设计, 注意句式长短结合, 以增强可读性。

Readability Analyzer (可读性分析器) 是一种用于提取英文文本基本可读性统计数据工具。它由贾云龙编程, 由徐佳进和贾云龙设计。此工具可以计算一些经典的可读性评分, 如 Flesch 阅读容易度(Reading Ease)和 Flesch-Kincaid 年级水平(Grade Level), 以及一些其他的词汇复杂性指标, 例如类/形符比(TTR)、标准化类/形符比(STTR)。还可以从 Results 中读取单词/标记、类型、词条、句子、平均词长(AWL)、平均句长(ASL)等描述性统计数据。

将《哈利·波特与魔法石》与《哈利·波特与死亡圣器》的文本分别导入 Readability Analyzer 运行 Analyze 功能, 得到如下分析结果:



Readability Analyzer 1.0 (Using Microsoft Office Word 199X)			
Settings	Results	About	
File	Reading Ease	Text Difficulty	Grade Level
cleaned J. ...	3.90	96.10	4.99

Figure 1. Readability analysis of harry potter and the philosopher's stone

图 1. 《哈利·波特与魔法石》可读性分析

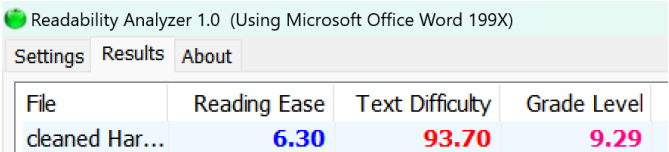


Figure 2. Readability analysis of harry potter and the deathly hallows
图 2. 《哈利·波特与死亡圣器》可读性分析

Table 10. Text readability score definitions
表 10. 文本可读性评分定义

Score mapping table		
Flesch Reading Ease Score		Readability Level
0~29	→	Very difficult
30~49	→	Difficult
50~59	→	Fairly difficult
60~69	→	Standard
70~79	→	Fairly easy
80~89	→	Easy
90~100	→	Very easy

Table 11. Text difficulty score definitions
表 11. 文本难度评分定义

Score mapping table		
Text Difficulty		Readability Level
0~29	→	Very easy
30~49	→	Easy
50~59	→	Fairly easy
60~69	→	Standard
70~79	→	Fairly difficult
80~89	→	Difficult
90~100	→	Very difficult

根据表 10 Flesch Reading Ease Score 和表 11 Text Difficulty 定义标准[6]可知,《哈利·波特与魔法石》和《哈利·波特与死亡圣器》的可读性均为非常难,并且后者较前者可读性较高,相对的,《哈利·波特与魔法石》和《哈利·波特与死亡圣器》的文本难度也非常高,且后者较前者文本难度更高。Flesch-Kincaid 年级水平评分将文本的可读性以美国学校年级水平来评分。例如,8.0 的得分意味着八年级学生可以理解该文档。对于大多数文档,目标是得分在 7.0 到 8.0 之间。而根据图 1、图 2 数据可得,《哈利·波特与魔法石》的 Flesch-Kincaid 年级水平评分为 4.99,意味着五年级左右的学生可以理解该小说,《哈利·波特与死亡圣器》的 Flesch-Kincaid 年级水平评分则为 9.29,意味着九、十年的学生才可以理解该小说。

3.2.2. 句子结构类型

英语句子按其结构可分为简单句、并列句和复合句。简单句只有一个主谓结构,并列句用并列连词

将两个或两个以上的简单句连在一起，复合句由从属连词将两个或两个以上简单句连接在一起。^[7]英语连词具有结构连接和语义连接两种主要功能。在故事撰写层面，使用连词可适当缩小阅读难度，让文章更连贯、更清晰。将经过词性赋码的两个语料子库导入 AntConc3.5.9，运行其 Word List 功能，检索并列连词(CC*CCB*)和从属连词(CS*CSA*CSN*CST*CSW*)的数量。

Table 12. Coordinating conjunction data of the two corpora
表 12. 两个语料子库的相关连词数据

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
从属连词	1022	1158
并列连词	2126	2792
总词数	77,916	205,920
从属连词占比	1.31%	0.56%
并列连词占比	2.73%	1.36%

表 12 数据显示，《哈利·波特与死亡圣器》文本使用从属连词和并列连词的占比低于《哈利·波特与魔法石》，表明《哈利·波特与魔法石》更善于使用并列句和复合句衔接文本，其信息单元之间的逻辑和语义关系更加紧密和清晰，也有利于读者更有效地理解句段和语篇。前文研究得出《哈利·波特与死亡圣器》的平均句长数值大于《哈利·波特与魔法石》的平均句长数值，但其句子内部架构却相对松散，逻辑不如《哈利·波特与魔法石》连贯，可读性相对较弱。

3.3. 语篇特征对比分析

文本长度在一定程度上可以反映篇章文体特征。一般而言，同一体裁，不同文本的篇幅可以说明信息负载的大小^[8]。

Table 13. Logical connective data of the two corpora
表 13. 两个语料子库的逻辑联系语数据

参数类型	哈利·波特与魔法石	哈利·波特与死亡圣器
从属连词	1022	1158
并列连词	2126	2792
连接副词	860	1245
总词数	77,916	205,920
逻辑联系语占比	5.14%	2.52%

根据表 13 可得：首先，从连词(包括从属连词、并列连词和连接副词)的使用差异来看，《哈利·波特与死亡圣器》的逻辑联系语占比低于《哈利·波特与魔法石》。这表明在构建句子和段落之间的逻辑关系时，《死亡圣器》使用了较少的显性连接手段。在文学作品中，连词是建立和维持篇章连贯性的重要工具，它们有助于读者理解文本中各个部分之间的逻辑关系。

其次，从句子结构的角度来看，《哈利·波特与死亡圣器》使用并列句、复合句和逻辑联系语的频率低于《哈利·波特与魔法石》。这进一步说明了《死亡圣器》在句子、段落和篇章层面的连贯性不及《魔法石》。并列句和复合句是构建复杂句子结构的重要手段，它们能够增加句子的信息密度和表达力，同时也有助于构建篇章的连贯性和紧凑性。

综合以上分析, 可以得出结论: 《哈利·波特与死亡圣器》在篇章的衔接性和紧凑性方面存在一定的不足。这是由于作者在创作过程中更注重情节的发展和角色的塑造, 而在篇章结构的构建上相对较为松散。当然, 这并不意味着《死亡圣器》的文学价值低于《魔法石》, 因为文学作品的评价涉及多个方面, 包括情节、人物、主题、语言等。但是, 从篇章结构和语言运用的角度来看, 《死亡圣器》在衔接性和紧凑性方面还有提升的空间。

4. 总结与讨论

本研究通过对 J. K. Rowling 的《哈利·波特与魔法石》和《哈利·波特与死亡圣器》两部作品的词汇、句式和语篇特征进行量化对比分析, 揭示了作者语言使用随时间推移的变化, 以及这些变化如何影响整体故事叙述的文体特征。

在词汇特征方面, 我们发现从《哈利·波特与魔法石》到《哈利·波特与死亡圣器》, 词汇长度、词汇密度、类符/形符比以及词目类型均有所变化。这些变化反映了作者随着故事发展和角色成长的需要, 对词汇选择的调整。例如, 随着故事的深入, 使用了更多复杂的词汇来描述复杂的人物情感和事件发展, 这体现在词汇长度的增加和词汇密度的提升上。此外, 高频词参数的变化也揭示了故事主题和焦点的转变。

在句式特征方面, 平均句长和句子结构类型的分析揭示了作者在句子构建上的变化。平均句长的增加表明作者为了更准确地传达情感、增强叙事节奏而选择了更长的句子。而句子结构类型的多样性则反映了作者为了保持故事的新鲜感和吸引力, 在句式上进行了更多的创新和尝试。

在语篇特征方面, 虽然本研究未提供具体的量化数据, 但可以推测, 随着故事从《哈利·波特与魔法石》的冒险起始阶段发展到《哈利·波特与死亡圣器》的高潮和结局阶段, 语篇结构、连贯性和叙事风格等方面也发生了显著变化。这些变化不仅体现在语言层面, 更体现在故事情节、角色塑造和主题深化等更宏观的叙事策略上。

综上所述, 我们需要注意到, 这些语言变化并不是孤立存在的, 而是与故事情节、角色成长和主题深化等叙事要素紧密相连[9]。因此, 在理解这些语言变化时, 我们需要将其置于整个故事框架中进行考量。此外, 虽然本研究采用了量化分析方法, 但语言使用的复杂性和多样性使得单纯的量化数据难以完全揭示其背后的深层含义。因此, 在未来的研究中, 可以进一步结合定性分析方法, 以更全面地揭示作者语言使用的变化和特点。

总之, 通过对《哈利·波特与魔法石》和《哈利·波特与死亡圣器》的对比分析, 不仅能够深入理解作者语言使用的变化和特点, 还能够更深入地理解整个故事框架和叙事策略。这对于提高翻译质量、推动文本分析研究的深入发展都具有重要意义。

参考文献

- [1] Scott, M. (2004) *The Wordsmith Tools*. Oxford University Press, 27-42.
- [2] Ure, J. (1971) Lexical Density and Register Differentiation. In: Perren, G.E. and Trim, J.L.M., Eds., *Applications of Linguistics*, Cambridge University Press, 443-452.
- [3] Butler, C. (1985) *Statistics in Linguistics*. Blackwell Publishers.
- [4] Laviosa, S. (2002) Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta*, **43**, 557-570. <https://doi.org/10.7202/003425ar>
- [5] 杨惠中. 语料库语言学导论[M]. 上海: 上海外语教育出版社, 2004: 42.
- [6] 刘靖, 黄立波. 《语料库文体学》述介[J]. 外语教学与研究, 2010(3): 236-239.
- [7] 王萃田. 英语的连词、连接和语义连接关系[J]. 外语学刊, 1985(1): 43-48.
- [8] 胡开宝, 谢丽欣. 基于语料库的译者风格研究: 内涵与路径[J]. 中国翻译, 2017, 38(2): 12-18+128.
- [9] 黄立波. 语料库译者风格研究反思[J]. 外语教学, 2018, 39(1): 77-81.