

# *The Stylistic Features of A Portrait of the Artist as a Young Man from the Perspective of Corpus Stylistics*

Yueyang Zhang<sup>1,a,\*</sup>

<sup>1</sup>College of Foreign Languages, China University of Geosciences, Xueyuan Road, Beijing, China  
a. 929343936@qq.com

\*corresponding author

**Abstract:** *A Portrait of the Artist as a Young Man* was Irish writer James Joyce's first published novel. The novel tells the story of the protagonist Stephen grow from a little child to an artist. Although Joyce uses third-person narration in the novel, he describes the surrounding society from Stephen's point of view everywhere. When Stephen grew up, his thoughts gradually matured, and he gradually changed from obedience to rebellion. The stylistic features of the novel also changed a lot in this process. Corpus stylistics is a scientific stylistic research method which combines qualitative research with quantitative research. With the support of corpus stylistics, the author studies and analyzes the stylistic changes in the novels during Stephen's childhood, adolescence and youth, and explains the causes of the corresponding characteristics. Through profound analysis, the author finally draws the conclusion: With the growth of the protagonist, the vocabulary of the novel becomes increasingly difficult and rich, the use of sentence grammar gradually becomes mature, and the difficulty of the text gradually increases. The style of the novel is closely related to the growth of the characters and the theme of each stage.

**Keywords:** a portrait of the artist as a young man, James Joyce, corpus stylistics

## 1. Introduction

### 1.1. Research Background

Due to the correspondence between corpus linguistics and literary research, since the 1980s, the academic circle began to explore the application of corpus in literary research, thus forming corpus stylistics [1]. Corpus stylistics aims to quantitatively describe and qualitatively analyze the stylistic features of various texts with the help of corpus tools by empirical means [2]. The core of this research method is to introduce data statistics into literature research, which can reduce the subjective arbitrariness of literature research.

*A Portrait of the Artist as a Young Man* is an autobiographical novel written by the Irish author James Joyce. The novel tells the story of the protagonist Stephen's growth from a babbling child to a young artist. Stephen experienced loneliness at church school, at home, in his faith and in love. He was ostracized and ridiculed by his schoolmates and was also confused by the politics and religion of his home country. As the story progresses, we can not only see the growth of the hero's age and the change of the space he lives in, but also the maturity of his mind and the growth of his talent from the change of style.

## 1.2. Purpose and Significance of the Research

This thesis hopes to creatively interpret the growth stages of the hero in *A Portrait of the Artist as a Young Man* by researching stylistic changes of the novel through the perspective of corpus stylistics.

First, through the use of corpus, this study can dig deeply into the stylistic characteristics of the work, which can help readers understand the connotation of *A Portrait of the Artist as a Young Man* and provide new ways of thinking for the work.

Second, this study uses corpus as a research tool to study literary works, which is of great significance to the development of corpus stylistics and helps to provide more cases for the application of corpus stylistics.

Third, innovatively studying *A Portrait of the Artist as a Young Man* with the corpus fills the gap in related research fields and helps to provide references for future related research.

## 2. Literature Review

Corpus stylistics is mainly applied to the study of English literary works in three fields. First, the study of the theme of literary works. Second, the study of the author's writing skills and literary techniques. Third, the study of the characters in the works.

### 2.1. The Study of the Theme of Literary Works

Corpus-based thematic studies of literary works often explore the themes of literary works by analyzing the subject words, high-frequency words or clusters of high-frequency words. Subject words refer to those words that appear more frequently in a text than in another, large-capacity corpus. Subject words or high frequency words usually reflect the topic or issue discussed in a work.

Stubbs did research and showed that the subject words of Conrad's novel *Heart of Darkness* are function words, abstract nouns and noun phrases with negative prefix adjectives. The application of these words highlights Marlowe's psychological uncertainty [3]. Based on relevant theories of functional linguistics, Liu Juan made a quantitative analysis of key words in Joyce's novel *Evelyn*, exploring Evelyn's mental journey during her escape and revealing the theme of the work [4].

### 2.2. The Study of the Author's Writing Skills and Literary Techniques

Corpus-based studies on writing skills and literary techniques often analyze the images and implications of literary works by analyzing high-frequency word clusters, so as to explore ways to advance the plots of literary works.

Mahlberg conducted a corpus investigation on 23 works of Dickens, analyzing high-frequency word clusters among them and finding that word clusters related to body parts were often clues to promote the development of key plots [5]. Some scholars found that in the English Gothic novel corpus, the occurrence frequency of time preposition word cluster and space preposition word cluster is much higher than that of other kinds of word cluster. Therefore, Gothic novels put more emphasis on the description of time and space when promoting plot development [6].

### 2.3. The Study of the Characters in the Works

Corpus-based research on character image usually analyzes the character's characteristics in literary works by searching the character's name or words referring to the character and analyzing the high-frequency words and keywords related to the character.

Inaki and Okita analyzed Alice's Adventures in Wonderland by using the corpus method and found that Alice was cautious and passive. While in Alice Through the Looking Glass, Alice is positive and optimistic [7]. Chen Chan conducted a corpus analysis of word clusters in 12 novels written by Alice

Munro, winner of the 2013 Nobel Prize in Literature, and found that the values of word clusters “as if” and those containing “she” were very high. Among them, “as if” is used in the description of the heroine’s appearance, behavior and psychological activities, which makes the description of female characters more vivid. The word clusters containing “she” are mostly combined with negative words, depicting the confusion and entanglement of the heroine [8].

### **3. Research Design**

#### **3.1. Research Questions**

This study is expected to summarize the stylistic characteristics of the novel at each stage of the protagonist’s growth by using corpus and tries to explain the similarities and differences.

#### **3.2. Research Method**

##### **3.2.1. The Combination of Qualitative and Quantitative Research**

This study uses corpus tools to conduct quantitative research on the text, which can present the stylistic features of the text more intuitively and scientifically, and will be supplemented by qualitative research to explain these features. The combination of the two research methods can give full play to the advantages of quantitative research and qualitative research while avoiding the disadvantages of both.

##### **3.2.2. Description, Interpretation and Evaluation**

It is not scientific to intuitively study the style of a work, but by describing and explaining some linguistic features of a text, readers can better understand the text. By evaluating these features, readers can better appreciate the aesthetic effects of the text. This is a stylistic analysis approach provided by Short: description, explanation, evaluation [9]. That is, first read the text, classify and describe the stylistic features of the work, then explain these stylistic features, and finally appreciate these features.

### **3.3. Research Procedure**

There are three main steps in this study. The first step is to divide the growth stage of the protagonist. The second step is to collect the electronic version of the novel and carry out noise reduction to obtain clean and accurate corpus. The third step is to use corpus software to calculate the required data and analyze the data.

#### **3.3.1. The Division of the Growth Stage of Stephen**

According to some literature and personal understanding of the novel, the author divides the growth process of Stephen into three stages. The first phase corresponds to the first chapter of the novel, which focuses on the childhood of the protagonist. In this phase, Joyce described Stephen’s life from babbling to primary school. The second phase corresponds to the second and third chapters of the novel, focusing on the protagonist’s teenage years. The third phase corresponds to the third and fourth chapters of the novel and focuses on the youth of the protagonist. In this phase, Stephen experiences his refusal to be a priest, his life in university and final determination to self-exile in order to pursue his artistic goal.

### 3.3.2. Corpus Construction

In order to compare the stylistic features of the three growth stages of the protagonist, we need to use corpus tools to calculate stylistic data of the three parts, respectively.

To construct the corpus of *A Portrait of the Artist as a Young Man*, the original work must be converted into a text format that can be read by each software. The author selected the original English text published by Yilin Press as the source material. The book is converted into txt format by scanning the paper version with the aid of OCR software and then rectifying the innumerable places of scanning errors. In the end, the author gets the clean txt document. The corpus has a size of 84922 words according to WordSmith.

### 3.3.3. Data Collection and Analysis

The average word length of a text refers to the average number of letters contained in all the words in a text, which can be abbreviated as AWL. To some extent, AWL can reflect the lexical difficulty of the text. If the AWL of one text is bigger, there are more long words in the text and the text is more difficult to read. The author uses the data calculated by Readability Analyzer. The software can also produce an estimated proportion of difficult words in the three parts.

Lexical density is the proportion of content words, which is a standard to measure the amount of text information. TTR is the acronym of type-token ratio, which refers to the ratio of types to tokens and can measure the lexical density of texts. Considering that type-token ratio will decrease with the increase of corpus volume, the author chooses the more reasonable standardized type/token ratio (STTR) for reference. The calculation method for STTR is as follows: Calculate the TTR of every 1000 words in each text, and average the various results obtained to get the mean value.

Word frequency can reflect the main idea of text narration to some extent. According to AntConc's word frequency retrieval function, the author finds that the word frequency of the three words "father", "god" and "i" has interesting changes in the three phases of the protagonist's growth.

The average sentence length (ASL) is one of the indexes which can reflect the difficulty of text to some extent at the syntactical level. Using Readability Analyzer, the average sentence length of Stephen's three growth stages can be calculated separately.

The Dale-Chall score is one of the most accurate readability metrics. Dale-Chall incorporates a list of 3,000 easy words which were understood by 80% of fourth-grade students. The readability score is then counted based on how many words in the passage are not in the list of easy words.

## 4. Research Results

### 4.1. Lexical Level

Table 1 shows the data calculated by Readability Analyzer. Both the average word length and the estimated proportion of difficult words indicate that the words in the first part of the novel are the easiest of the three parts, while the words in the last part are the most difficult of the three parts. The three sections are progressively difficult in terms of vocabulary.

Table 1: AWL and proportion of difficult words calculated by readability analyzer.

Growth Stage of Stephen	Average Word Length	Percentage of Difficult Words
The First Phase	4.08	4.17%
The Second Phase	4.30	7.38%
The Third Phase	4.33	8.03%

Table 2 shows the STTR of three phases of Stephen’s growth calculated by WordSmith. The STTR increases from the first part of the novel to the third part, indicating that the first part of the novel has the lowest lexical richness, while the last part has the highest lexical richness, suggesting the gradual maturity of Stephen's language and thoughts.

Table 2: Tokens, types and STTR of three phases calculated by WordSmith.

Growth Stage of Stephen	Tokens Used for Word List	Types	STTR
The First Phase	17719	2409	36.49
The Second Phase	31227	5163	43.86
The Third Phase	35951	5853	44.19

According to AntConc’s word frequency retrieval function, the author finds that the word frequency of the three words “father”, “god” and “i” has interesting changes in the three phases of the protagonist's growth. Table 3 shows the detailed data. In the first stage, Stephen in his childhood relied on his family and father a lot, so the frequency of using “father” is very high. However, as Stephen grew up, he was gradually disappointed in his family, and the word “father” appeared less frequently in the later two phases. In the second stage, Stephen was trapped in the cage of desire. After waking up, he repented for his fall and placed his soul in religion, so “god” appeared more frequently in the second phase than in other phases. In the third stage, Stephen gradually matures, awakening his self-consciousness, beginning to rebel against conventional life. The frequency of “i” increases significantly, while the frequency of “father” and “god” decreases correspondingly.

Table 3: Word frequency and rank of “father”, “god” and “i” calculated by AntConc.

Word	Rank in Phase1	Word Frequency in Phase1	Rank in Phase2	Word Frequency in Phase2	Rank in Phase3	Word Frequency in Phase3
father	32	81	49	73	156	28
god	63	37	25	143	94	43
i	26	96	29	130	11	347

## 4.2. Syntactical Level

Using Readability Analyzer, the average sentence length of Stephen's three growth stages was calculated separately and presented in Table 4. According to the table, the average sentence length in the first phase of Stephen’s growth is the shortest, which means the sentence structure of the first phase is relatively simple. The second phase has the longest average sentence length, which means the sentence structure of the second phase is more complicated. After manual analysis of the novel, the author found that the reason for above data was that Joyce uses lots of simple and incomplete sentences in the first stage of Stephen's growth to show the thinking and language characteristics of a child. In the second stage, young Stephen’s language system developed well, often using a more standardized grammar of long sentences. In the third stage, young Stephen likes to resort his thoughts to long poems, and the sentence structure tends to be concise and mature.

Table 4: ASL of Stephen's three growth phases calculated by readability analyzer.

Growth Stage of Stephen	Average Sentence Length
The First Phase	11.59
The Second Phase	17.44
The Third Phase	14.52

### 4.3. Textual Level

By using Readability Analyzer, the Dale-Chall scores of Stephen's three growth phases are calculated separately and shown in Table 5. The score increases from the first stage to the third stage, indicating that the reading difficulty of the novel text also increases accordingly, and Stephen's language and thought are gradually mature and formal.

Table 5: Dale-Chall Scores of Stephen's three growth phases calculated by readability analyzer.

Growth Stage of Stephen	Dale-Chall Score
The First Phase	6.13
The Second Phase	7.18
The Third Phase	7.32

## 5. Conclusions

In order to study the stylistic characteristics of the novel, the author uses the corpus software to calculate the average word length, the proportion of difficult words, the lexical richness, the high-frequency words, the average sentence length and the text difficulty coefficient of the hero in the three growth stages of the novel. After obtaining the corresponding data, the author analyzes and interprets the data, and compares the data between the three phases.

Finally, the author draws the following conclusions: With the growth of the protagonist, the vocabulary of the novel becomes increasingly difficult and rich, the use of grammar gradually becomes mature, and the difficulty of the text gradually increases. The different high-frequency words in the three stages also show that Stephen's beliefs are constantly changing and he finally returns to his true heart to find himself. The style of the novel is closely related to the growth of the characters and the theme of each stage.

## References

- [1] Hu Kaibao and Yang Feng (2019) *Corpus-Based Literature Research: Connotation and Significance*. *Journal of Zhejiang University (Humanities and Social Sciences Edition)*, 5, 143-156.
- [2] Liu Jing and Huang Libo (2010) *Introduction to Corpus Stylistics*. *Foreign Language Teaching and Research*, 3, 236-239.
- [3] Stubbs, M. (2005) *Conrad in the Computer: Examples of Quantitative Stylistic Methods*. *Language and Literature*, 14(1), 5-24.
- [4] Liu Juan (2013). *The Blending of Morality and Emotion: Evelyn's Spiritual Flight from the Perspective of Corpus Stylistics*. *Foreign Languages*, 4, 48-52.
- [5] Mahlberg, M. (2007) *Clusters, Key Clusters and Local Textual Functions in Dickens*. *Corpora*, 2(1), 1-31.
- [6] Ren Yan, Chen Jiansheng and Ding Jun (2013) *Word Clusters in British Gothic Novels: A Corp-Based Stylistic Study*. *Journal of the PLA Institute of Foreign Languages*, 5, 16-20+127.
- [7] Inaki, A. and Okita, T. (2006) *A Small Corpus-Based Approach to Alice's Roles*. *Literary and Linguistics Computing*, 21(3), 283-294.
- [8] Chen Chan (2014) *The Features and Functions of Word Clusters in Alice Munro's Novels: A Stylistic Study Based on Corpora*. *Journal of the PLA Institute of Foreign Languages*, 3, 151-159.
- [9] Short, M. (1997) *Exploring the Language of Poems, Plays and Prose*. Harlow: Addison Wesley Longman Limited.