

A Corpus-based Study of James Joyce's *A Portrait of the Artist as a Young Man*

Muhammad Ajmal
University of Lahore

Aisha Shoukat
University of Lahore

Abstract

Exploring literary text(s) with corpus-based approaches is becoming popular and enlarging our understanding of a particular phenomenon in a text. This study aims to encourage novice researchers to utilize corpus-based approaches while conducting research in the domain of literature. The present study applies a corpus-driven approach to James Joyce's *A Portrait of the Artist as a Young Man*, to see how well this new method works with literary texts. It has been found that the corpus-driven approach can provide quite a few results that are useful in supporting and refining literary scholars' intuitive observations on the author's works. The significance of the title and concordance hits and concordance plot of the words *portrait*, *artist*, and *young man* in the title are observed. The imperative sentences used are explored throughout the novel. Out of all linguistic features, thirty most used words were collected, which belonged to different parts of speech and their frequency were noted. The analyst's knowledge and understanding of the text(s) under study are crucial in interpreting and evaluating those results because the corpus-driven approach to literary texts relies heavily on quantitative data. The study affirms that the text can be well interpreted and explained by using AntConc software.

Keywords: AntConc 3.5.8., concordance, corpus-based analysis, corpus linguistics, literary text

1. Introduction

Corpus can be defined as a collection of written texts which include the entire work of an author or a piece of writing on a particular subject. The methods used in Corpus Linguistics were adopted in the 1960s but until the 1980s the term corpus linguistics did not exist. It is a methodology that is used quantitatively or qualitatively to secure and study the language data. Computers are a necessary tool used for this purpose. This approach can be applied in any area of language studies and support the empirical investigation of language variation. While a corpus linguistic technique has been applied to various studies in text and discourse analysis, it has not been much adopted in stylistic analysis of literary texts. It has been found that the corpus-driven approach can provide quite a few results that are useful in supporting and refining literary scholars' intuitive observations on the author's works. Some of the linguistic patterns derived from the comparative corpus-driven method have not been remarked on before in any previous studies. Despite such great potential for the study of style in literary works, it is suggested that the analyst's knowledge and understanding of the text(s) under study are crucial in interpreting and evaluating those results because of the corpus-driven approach to literary texts relies heavily on quantitative data. AntConc 3.5.8 developed by Laurence Anthony was utilized to get quantitative results. The words in the title were analyzed first. The number of times the words: *portrait*, *artist*, *young man*, appeared in the novel were noted. Later on, the thirty most frequently used words were searched and were noted with their frequency. The words belonged to a different part of speech.

Imperative sentences are said to be the type of sentences that instruct someone or give advice, express command, or give an order, tell the direction to someone or make a request. They can be long as well as short that depends on their context. It has a significant role in writing as well as speaking, also used for instructions on signs on the road and in advertisements. Writers use imperative sentences to give a clear idea, to express emotions according to the ongoing situation between the characters, likeness love, or hate feelings through their writing.

The story of the *A Portrait of the Artist as a Young Man* revolves around the Character Stephen Dedalus, he shows the religious and intellectual awakening. The hero of the novel Stephen is a sagacious third-person narrator, but the free indirect style is adopted by Joyce. The main feature used by Joyce is the use of a stream of consciousness. This technique makes the novel a story of the development of Stephen's mind. The hero of the novel Stephen thinks differently. He is unable to balance art and religion. He hears the talk between his father and Conmee regarding one incident and he realizes that he was kept in dark. He thinks that he and his father are no more friends and ponders why no action was taken against his father. His classmates who created trouble for him mocked him. He learned things quickly; especially his anger helped him a lot. He was mature enough at an early age and become an artist. After his commitment to sin with a prostitute, he hears sermons and repents on sins. He realizes that sin is a sin whether it gives pleasure or not. He realizes that the actual meaning of the uttered word is what depends on the mind of the listener or reader. He preferred art over religion and his family. Stephen is able to describe his world in simple words and phrases. The opening section of the novel progresses logically. Stephen observes the situation. He becomes aware of his surroundings. Stephen is a depiction of Joyce himself.

2. Literature Review

From the past few decades, we have witnessed a drastic and prominent application of corpus-based techniques in the field of applied linguistics to highlight the usage of linguistic devices in both the literary and non-literary text. It helps the analysts and researchers to come up with the exact numbers and co-occurrences of linguistic elements, which gives individuality to any writer's work. It provides the researchers with evidence to prove the different usage and diversified occurrences of linguistic elements. For this purpose, it is more considered to be a scientific approach as the results are quantifiable and thus are more reliable. Some corpus linguistics like Mahlberg (2005) and Bonelli (2001) believe that this kind of corpus-based research allows the analysts to test their hypothesis about certain linguistic or textual features and the results are obtained in the form of quantitative data. This quantifiable data can then further be explained semantically or pragmatically thus figuring out the style of the text. The use of corpus linguistics, or in particular its research techniques, has also been extended to literary analysis. These techniques, either quantitative or qualitative or both, can result in very detailed stylistic descriptions of literary texts that can either complement conventional interpretations or enable insights not possible through intuition alone.

Stubbs and Halle (2012) define Corpus Linguistics as it is a computer-assisted method to deal with the larger quantity of linguistics analysis in a text of novels or anthologies. It is used to check the validity of language produced by the non-expert users of language that have started to write in newly learned international languages. Sinclair (1997) describes that Corpus Linguistics is the study of language structures, but it is different from customary linguistics. Its focus is on the

systematic study of the authenticity of language in use. He claimed that corpus linguistics does not define just a newly emerging methodology for studying language, but also a new research initiative and a new philosophical approach to the subject. Hockey (2001) says that the wordlist tool of AntConc 3.5.8 software generates words in a corpus with their frequency order. It also provides the ability to count the words according to their “stem” form. Lonfils & VanParys (2001) while discussing AntConc 3.5.8 software expressed that avoidance of pull-down menus and additional windows it's an important feature of good software. Hunston (2002) says that corpora have been used in a wide range which includes areas related to translation studies, stylistics, grammar, and dictionary.

Corpus stylistics is a branch of computational linguistics as Wales (1989) points out. It was developed in the late 1960s. It helps to investigate certain characteristics of the data like the length of words and sentences based on statistical and computer-aided tools to study several issues related to style (85). One important issue needs to be given attention to here in the study, McIntyre (2015) sheds the light on the difference between corpus stylistics and corpus linguistics. He asserts that many works have developed the sub-branch of stylistics namely, corpus stylistics, but none of them give an accurate definition of corpus stylistics and differentiate it from corpus linguistics (60). Then, he defines corpus stylistics, as “corpus stylistics is simply corpus linguistics with a different object of study (literature as opposed to non-literary language). Besides, he demonstrates that the difference between them is that corpus stylistics is not only borrowing tools from corpus linguistics but it makes itself unique by using qualitative tools and techniques of stylistics to analyze texts with the help of computational methods. Semino and Short (2004) assert that doing corpus-based work is highly quantitative, but this matter of fact does not mean that qualitative analysis should be excluded. On the contrary, if both are combined they help to reach a higher degree of understanding of literary works or any other selected data.

Noguchi (2004) stresses the importance of AntConc 3.5.8 and suggests an easy way to use toolkit for corpus analysis. It is lightweight, simple, and easy to use. Moreover, it is extremely effective to use in the classroom. According to Mahlberg corpus linguistics can be defined as an approach to the description of English with its theoretical framework. According to McEnery’s point of view, corpus linguistics can be termed as a complete system of method and principles about how to apply corpora in language teaching, learning, and language studies. It has a theoretical status and it is not a theory in itself. Thompson and Hunston (2006) quoted that corpus linguistics is a methodology and can be associated with any theoretical approach to language. Kuebler and Zinsmeister answered a question in “Corpus Linguistically Annotated Corpora” (2015) about corpus linguistics they said that corpus linguistics can be either a theory or a tool. It can be determined by how it is applied. This study endeavors to find out answers to the following research questions:

1. How do the results encourage novice researchers of literature to utilize corpus-based studies while conducting research?
2. What are the most frequently used words in *A Portrait of the Artist as a Young Man*?

3. Research Methodology

This study utilizes Mahlberg & McIntyre's (2012) methodology which focuses on one single literary text by one author. They explain that studying one text may be considered as a "small sample of data" but they assert that this text is still regarded as part of a corpus. The nature of research is both qualitative and quantitative. The Internet was the main source by which data was collected. After going through different articles and observation of the suitable software for the analysis AntConc software was used to analyze the text. The importance of the title including the concordance plot of the words "portrait", "artist" and "young man" in the title "*A Portrait of the Artist as a Young Man*" and the number of times the words appeared in the novel were noted. The imperative sentence in the novel was observed. Out of all linguistic features used in the novel, a word list of thirty most frequently used words by the author was created along with its frequency for which no specific category of part of speech was specified so objectivity is reached. The novel was analyzed in a limited time. The study showed that it is easier to obtain data by computer than manually. It examines the significance of acquiring and analyzing data in less time. The screenshot and charts are provided where required.

4. Results

The corpus-based analysis of *A Portrait of the Artist as a Young Man* utilized AntConc3.5.8 (windows) 2019. The research questions were explored and results were provided for a better understanding. Significant importance of the title, the number of times the words "portrait", "artist" and "young man" appeared in the novel, and their concordance plot in the title *A Portrait of the Artist as a Young Man*. The title is able to show its open-ended portrayal of the whole story as penned by Joyce. The main theme involves the development of the artist. And as the choice of words by Joyce, the selection of articles to show the actual meaning of the words chosen by Joyce for the title. The article "a" before portrait shows that Joyce is not considering many portraits its only one portrait or only one side shown by Joyce. The article "the" before artist gives it an autobiographical touch as well as indicates a form of art that he wants to create on his own. Stephen Dedalus can be a portrait of Joyce himself but according to Joyce's biography, the character of Stephen differs a bit from the beginning of the story. When we reach the 5th chapter we come to know that the word "artist" points to a literary artist. The words "artist" and "portrait" raise Stephen's identity as an artist. Stephen is an aspiring writer. The word "portrait" shows his love for art and his ambition to create a representation of reality. The ending words of the title "young man" express Stephen's inexperience and growing pains. The title shows the greatest coming of age stories in English Literature. The word "portrait" is a strange choice, as a portrait is a static object which cannot move. The story of the novel starts from Stephen's childhood and ends up Stephen as a young artist setting himself off into the world to pursue his art. Joyce added a static form to the novel and also show the development and evolution of Stephen himself.

The words used in the title of the novel were analyzed using AntConc software. By analyzing the words in the software the results that were found are given in the Figure 1: Analysis of words in the title of the novel "*A Portrait of the Artist as a Young Man*"

The word “Portrait” is used 160 times in the novel, “artist” 171 times, and “young man” 163 times. The artist is used more as it was the end of the novel. The results can also be shown using a bar graph as given below:

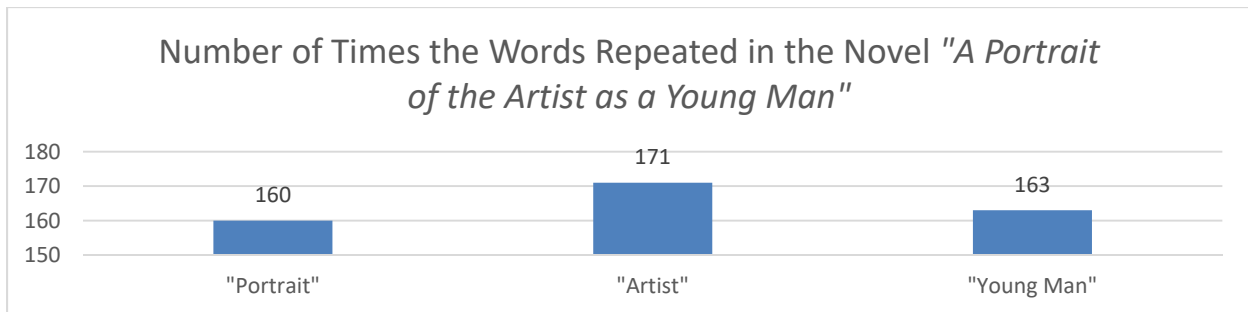


Figure 1: Analysis of word: *portrait*, *artist* and *young man*

A concordance plot is can be defined as a plot that shows where the search word appears in a set of text. It shows a very simple visualization of Key Word in Context (KWIC). The instance where the word is present in the novel is represented with a black line. Following are the concordance plots of the words “portrait”, “artist”, and “young man” in the title “A Portrait of the Artist as a Young Man”.

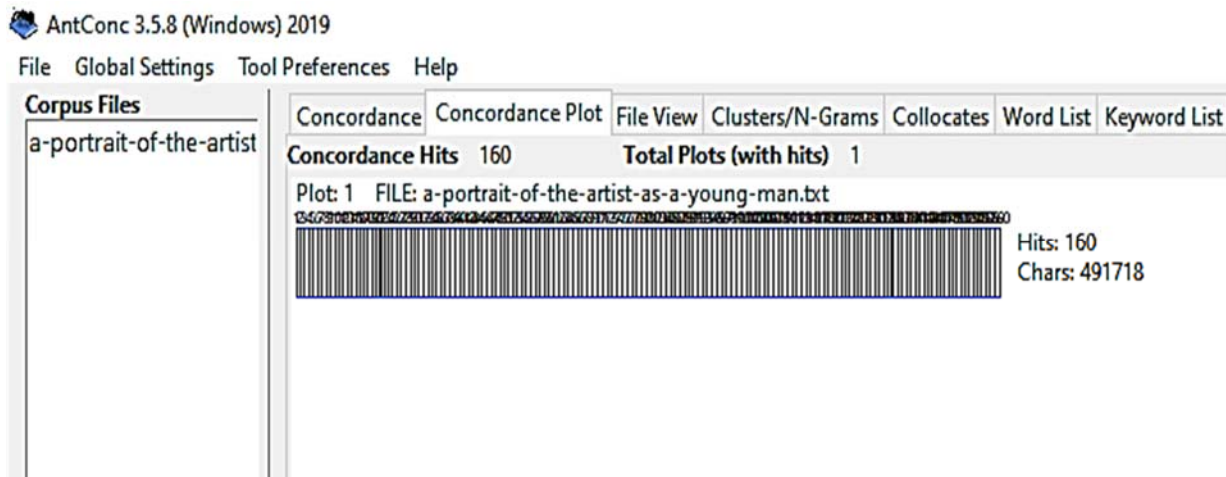


Figure 2: Concordance plot of the word *portrait*

The Figure 2 shows the concordance plot of the word “portrait”. It shows the hits 160 times. It means Joyce used the word in the novel 160 times. The black lines in the bar show the instances where the word is present in the novel. The chart gives the length of the file. That is 491718 for the word “portrait”.

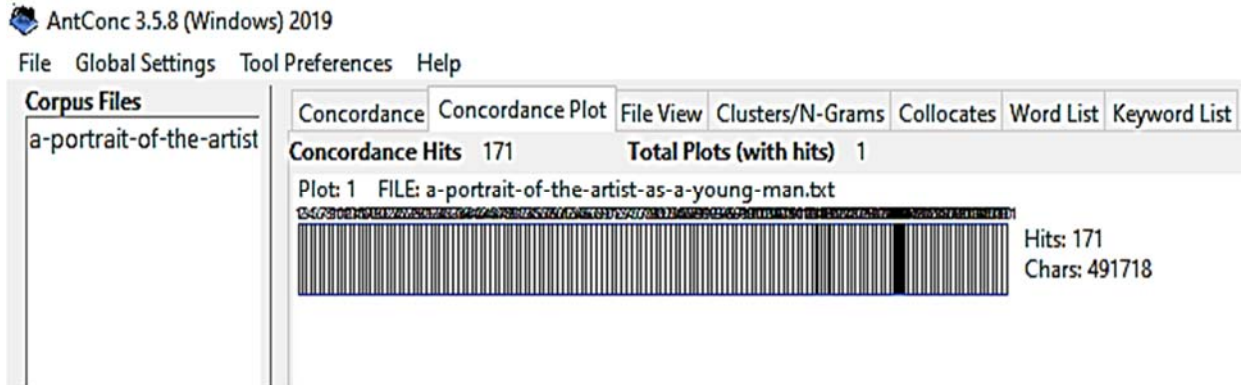


Figure 3: Concordance plot of the word *artist*

Figure 3 shows the concordance plot of the word “artist”. It shows the hits 171 times. It means Joyce used the word in the novel 171 times. The black line in the bar shows the instances where the word is present in the novel. The chart for the word “artist” is the same as the Chart for “portrait” that is 491718.

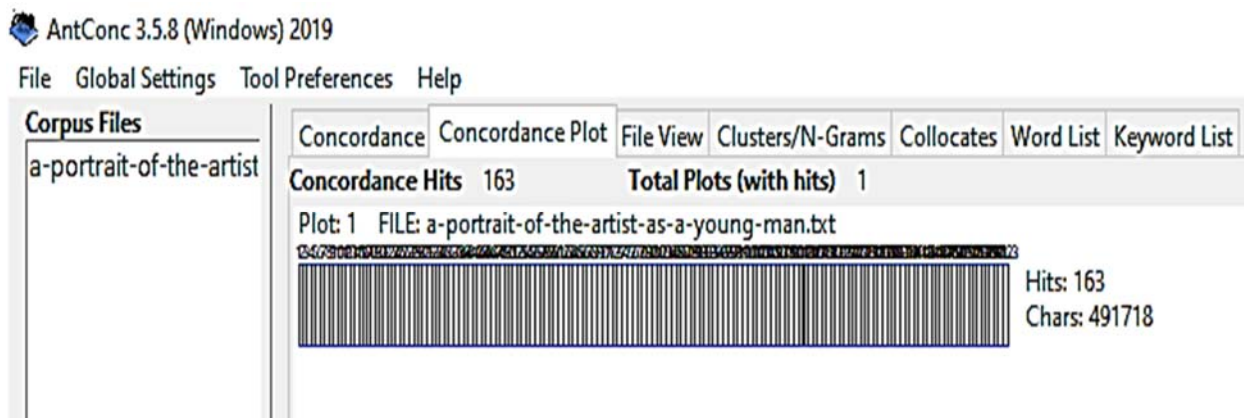


Figure 4: Concordance plot of the word *young man*

Figure 4 shows the concordance plot of the word “young man”. It shows the hits 163 times. It means Joyce used the word in the novel 163 times. The black line in the bar shows the instances where the word is present in the novel. Chart for the word “young man” is the same as the Chart for “portrait” and “artist” which is 491718.

An imperative sentence can be defined as a type of sentence that instructs or give advice, express command, give an order, tell a direction, or make a request. Depending upon its context it could be simple and short, or long and complex. It plays an important role in writing and speaking as well. It is important in everyday conversational language rather than in literary writing. It is commonly used in signs on the road, to give instructions, in advertisements. It requires a specific audience to deliver. Writers use imperative sentences to give a clear and straightforward command and instructions, to express feelings and emotions of displeasure, likeness, love, or hate through their writing. .The imperative sentences used in the novel “*A Portrait of the Artist as a Young Man*” by James Joyce were found using Lawrence Anthony’s software, the AntConc software. The following table elaborates on the words beginning with the imperative sentences and how many times the word has been used throughout the novel.

Table 1: words used as imperative

Sr. no.	Imperative	Appearance in text
1	Remember	39
2	Read	31
3	Try to	15
4	Bring	8
5	Please	7
6	Let him	6
7	Give me	5
8	Forgive	4
9	Do it	3
10	Help you	3
11	Hurry up	3
12	Say it	3
13	Get out	2
14	Take it	1
15	Sit down	1

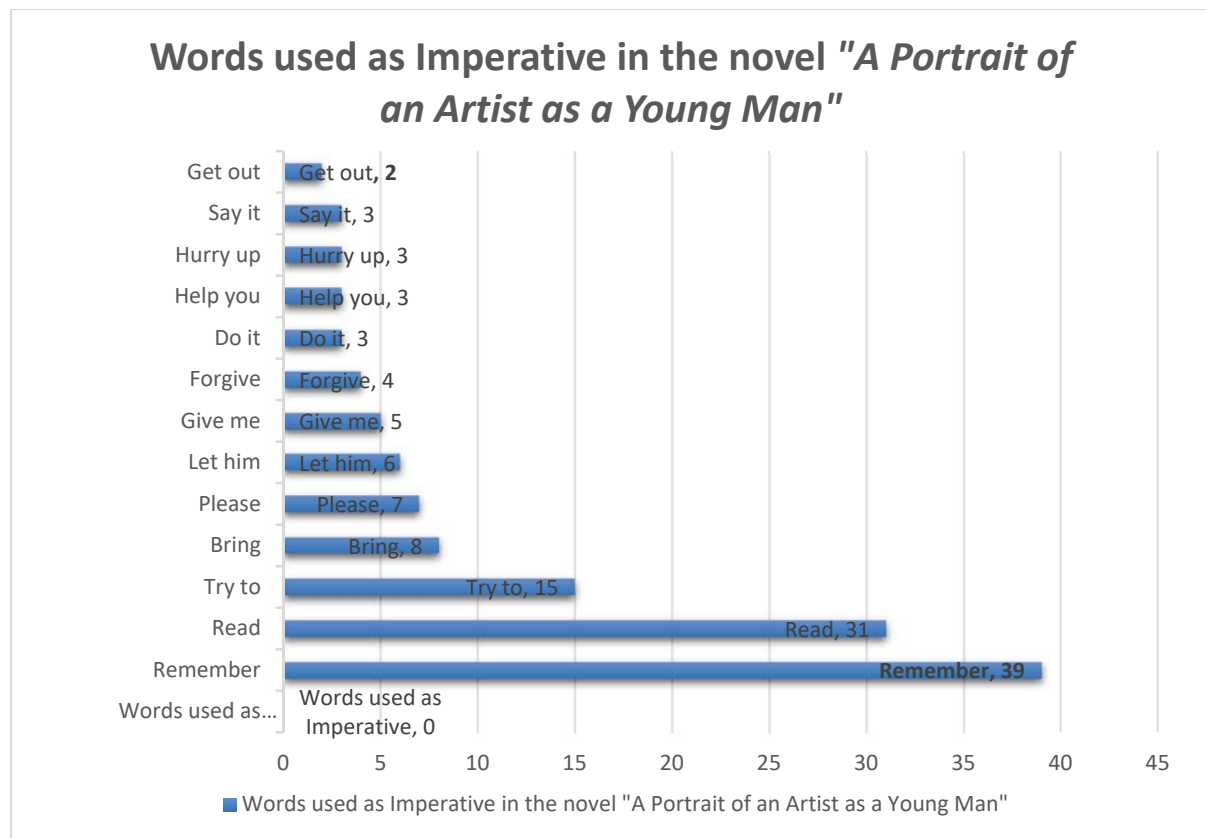


Figure 5: Analysis of the words used as Imperative

The bar chart shows the usage of different words with the number of times the word appeared in the novel "A Portrait of the Artist as a Young Man". Joyce has used his linguistic featured very skilfully. He has used imperative expressions throughout the novel. He created a connection between the character and the imperative sentences that made them connected. He has

presented clear ideas and the language of the Novel very well portrays emotion in the novel and it. The software has also helped to extract every sentence by simple one click.

The novel “*A Portrait of an Artist as a Young Man*” was analyzed using the AntConc software. Wordlist means to create a list of all the words in the corpus. The total words generated in the corpus are 9528. Given below is the screenshot taken from the software, which includes the first fifteen most frequent words with their frequency from the complete novel.

Corpus Files		Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
a-portrait-of-the-artist		Word Types: 9528	Word Tokens: 89248	Search Hits: 0				
Rank	Freq	Word						
1	6070	the						
2	3375	and						
3	3308	of						
4	2266	a						
5	1937	to						
6	1878	he						
7	1745	his						
8	1557	in						
9	1066	was						
10	993	that						
11	902	had						
12	868	it						
13	722	you						
14	706	him						
15	644	at						

Figure 6: Wordlist of frequently used words

Figure 6 shows the words which are frequently used in the novel. The words do not belong to one part of the speech. The words belong to different parts of speech like “article”, “conjunction”, “preposition”, “pronoun”, and “verb”. The frequency shows how many times the word has been used. The article “the” a is used 6070, 2266 times in the novel. Conjunction “and” is used 3375 times. Preposition “of”, “to”, “in”, “it”, “at” is used 3308, 1937, 1557, 868, 644 times. Pronoun “he”, “his”, “you”, “him” is used 1878, 1745, 722, 706 times. Verb “was”, “had” is used 1066, 902 times. The demonstrative adjective *that* is used 993 times. The next figure no 5 below shows the next 15 words used frequently in the novel.

AntConc 3.5.8 (Windows) 2019
 File Global Settings Tool Preferences Help

Corpus Files
 a-portrait-of-the-artist

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Word Types: 9528 Word Tokens: 89248 Search Hits: 0

Rank	Freq	Word
16	629	as
17	615	said
18	601	i
19	573	with
20	538	on
21	536	for
22	493	s
23	483	from
24	449	not
25	447	is
26	421	stephen
27	401	by
28	387	they
29	383	be
30	356	but

Figure 7: Wordlist of frequently used words

Figure 7 shows the words that are frequently used in the novel. Like, figure 4 the words here too belong to a different part of speech. The frequency of the words in the figure are: Conjunction “as”, “but” is used 629, 356 times. The preposition “with”, “on”, “for”, “from”, “by” is used 573, 538, 536, 483, 401 times. The pronoun “I”, “they” is used 601, 387, times. The verb “said”, “s”, “is” “be” is used 615, 493, 447, 383 times. The adverb “Not” is used 449 times. Noun “Stephen” is used 421 times.

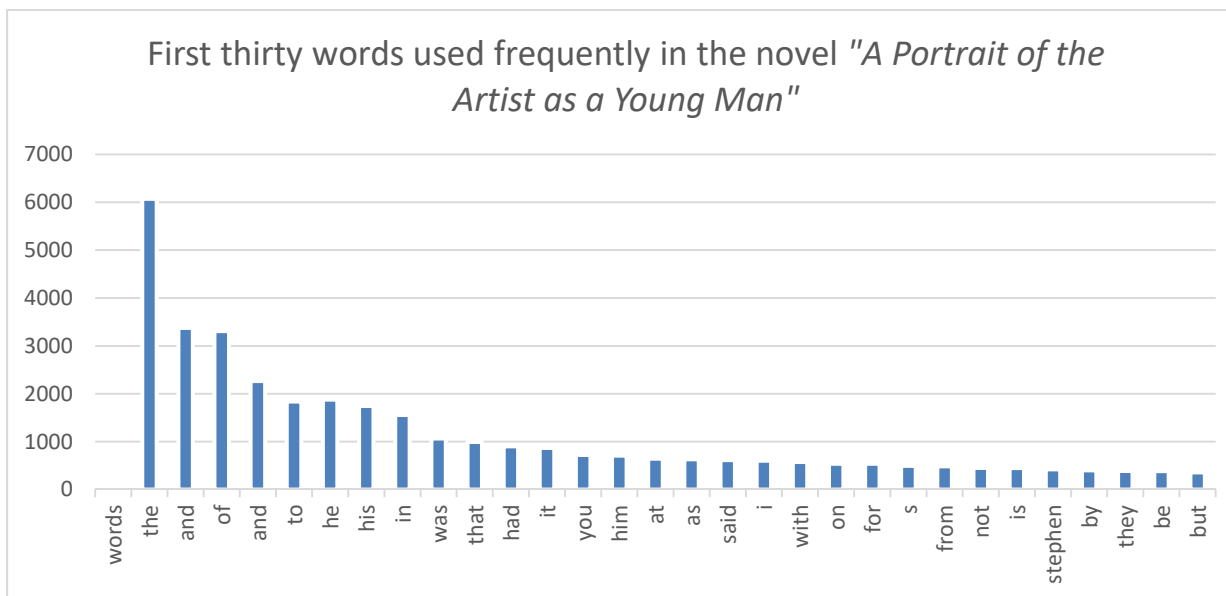


Figure 8: Most frequently used words

The results above show the first thirty frequently used words in the novel “A Portrait of the Artist as a Young Man” the words do not belong to a single part of speech. They belong to different parts of speech like noun, verb, adverb, conjunction, articles, etc.

5. Discussion

The results reveal the frequency of words used as imperative in the novel. The word “remember” has been repeated 39 times in the novel. The phrases such as “remember all”, “remember it”, “remember only”, “remember” all such phrases that depict that the characters are indicated to remember something. The recalling of memories and remembering or asking to remember to show a close relationship between the characters of the novel jot down by Joyce.

The word “read” has been repeated 31 times in the novel. The word “read” used as a phrase like “read a line”, “read above”, “read down”, “read from”, “read for”, “read it”, “read on”, “read the flyover”, “read three reviews”, “read without”. It shows that the characters order each other to read something. Alternatively, they command each other to read from a particular place. On the other hand, read a specific number of pages or reviews. Joyce has wonderfully used each phrase and tried to convey emotions and desires among his characters and indulge the readers in reading the novel.

The word “Try to” has been repeated 15 times in the novel. The phrases in a sentence or the complete sentences that contain the word “try to” in it shows some indication of advising to do something. The phrases used in the novel are “try to catch”, “try to save her”, “try to learn”, “try to induce”, “try to understand”, “try to trim” all indicate such actions that need some advice from one character of the story to the other.

The word “bring” is used in the novel as “bring”, “bring him”, “bring in”, “bring it”, and “bring us”. The characters request to bring something either physical, to recall in memories, or to bring vegetables. The characters are connected throughout the novel.

“Please” has been used 7 times in the novel. In the sentences, the words “please come and take me home”, “please copy”, “please God”, “please remember”, “please sir”, “please teacher”, “please” all the words express some request. The writer intends to express the feeling of request from the characters. The character requests each other to do a certain task.

“Let him” has been used 6 times in the novel. The phrases used by Joyce in the novel with “let him” are, “let him remember”, “let him in”, “let him see”, “let him stick” all such phrases show that one character waits for the other to do a certain action or to help them to do a certain task. The pronoun “him” shows us that the task of the order belongs to the male character of the novel.

The word “give me” has been repeated 5 times in the novel. The characters are asking for something to be given. By reading the sentences, we come to know what the characters are asking for. Some ask for another cigarette, some say to give them paper, some others ask for providing them with hypotenuse. The word chosen by the writer well expresses emotions that the character wants to convey to each other. Joyce has conveyed the feelings and emotions using his narrative techniques, which every character has expressed in his or her dialogue and can be well interpreted by the readers that help them to understand the story well.

“Forgive” has been repeated four times in the novel. The expressions Joyce used in the novel with the word “forgive” include, “forgive him”, “forgive me”, and “forgive you”. Joyce has tried to show the kind act of forgiving people and asking forgiveness from God both acts help to comfort mind and soul. It relaxes our mind. 3 times word repeated word “do it” in the sentences used by Joyce gives the command to do something.

“Help you” has been repeated 3 times in the novel. As the word suggests to help likewise Joyce has expressed the feeling of advice from a character to another that they will help or asked one another so they can help each other. “Hurry up” was used 3 times in the novel. The words express that something is going on and one character says to another to hurry up and finish it. The word “say it” repeated 3 times in the novel shows the feeling of anger and also indicates that a character asks another to say something loud.

The word “get out” has been repeated 2 times in the novel. It orders the character to get out from somewhere. Joyce has tried to express every emotion and desire and the language used to make it even more expressive. “Take it” has been used one time in the novel. It shows a feeling of calm and cool mindedness. It is an advice from one character to another character to take the thing easy. “Sit down” used one time in the novel shows a command to sit down and eat his dinner.

6. Conclusion

The study highlighted Corpus Linguistics as a powerful tool in the analysis of literary texts, especially when allied with non-computational approaches, to bring into light interpretations, thematic details, critically important words in a text, and other information that to other types of analysis might go unseen. By enabling the researcher to process a large quantity of data, and by giving a statistical treatment to the information obtained, Corpus Linguistics provides an ideal approach to study various characteristics of a literary text that would otherwise have gone unnoticed. The study concludes by showing the significance of using a corpus-based analysis of the literary text (s). *A Portrait of the Artist as a Young Man* by James Joyce was analyzed using AntConc software. The number of times the words “portrait”, “artist” and “young man” used in the novel with their concordance plot were explained. The imperative sentences used in the novel were all observed and the numbered their occurrence in the table. Then the most frequently used words in the novel were investigated. The results indicate that these words belong to different parts of speech. The study provides a new platform for future researchers to explore new dimensions in the domain of literature by using corpus software because the number of words used in sentences can be easily extracted. Through the AntConc 3.5.8 software tools, the researchers may identify the frequency, wordlist, concordance, and concordance plot. To conclude, corpus stylistics is incredible assistance in directing such investigation.

References

- Anthony, L. (2020). *Lawrence Anthony's Website*. Retrieved from Lawrence Anthony's Website: <https://www.laurenceanthony.net/software/antconc/>
- Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Kazemian, B., & Hashemi, S. (2017). A Radical Shift to a Profound and Rigorous Investigation in Political Discourse: An Integrated Approach. *International Journal of English Linguistics*, 7(3), 115-128. <https://doi.org/10.5539/ijel.v7n3p115>
- Kuebler, Zinsmeister. (2015). *Corpus Linguistics and Linguistically Annotated Corpora*.
- Lakoff, G. & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.
- Lennard, J. (1996). *The Poetry Handbook*. Camb Oxford University Press.
- Leech, Geoffrey. (1992). *Corpora and theories of linguistic performance*. In J.Svartvik (ed.), 105–122.
- Thompson, Geoffrey, and Susan Hunston (Eds.). 2006. *System and corpus: Exploring connections*. London: Equinox.
- Lewis, P. (2007). Cambridge Introduction to Modernism (Cambridge UP, 2007), pp. 129-151.
- Lonfils, C., and Vanparys, J. (2001). How to design user-friendly- CALL interfaces. *Computer Assisted Language Learning*, 14(5), 405-417.
- Mahlberg, Michaela. (2005). *English general nouns: A corpus theoretical approach*. Amsterdam: John Benjamins.
- McEnery, Tony, Richard Z. Xiao, and Yukio Tono. (2005). *Corpus-based language studies: An advanced resource book*. London: Routledge.
- McIntyre, D. (2015). *Towards an Integrated Corpus Stylistics*." *Topics in Linguistics*: 59-68.
- Newmark, P. (1981). *Approaches to Translation*. New York: Pergamum Press.
- Newmark, P. (1988). *A Textbook of Translation*. London: Prentice hall International (UK) Ltd.
- Noguchi, J. (2004). *A genre analysis and mini-corpora approach to support professional writing by non-native English Speakers*. *English Corpus Studies*, 11, 101-110
- Ross, Andrew. (1984). *The Waste Land and the Fantasy of Interpretation*." *Representations* 134-58. JSTOR. Web. 5 Feb. 2012.
- Semino, Elena, and Mick Short. (2004). *Corpus stylistics: Speech, Writing, and Thought presentation in a corpus of English writing*. London: Routledge, Print
- Steen, G. J. et al. (2010). *Metaphor in usage*, *Cognitive Linguistics*, 21 (4).
- Stephen, M. (1984). *An Introductory Guide to English Literature*. Hong Kong: York Press.
- Thompson, Geoffrey, and Susan Hunston (Eds.). (2006). *System and corpus: Exploring connections*. London: Equinox.
- TogniniBonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: John Benjamins.
- Wales, Katie. (2014). *A Dictionary of Stylistics*. London: Routledge, Print.