



A Corpus Stylistic Analysis of Linguistic Differences in James Joyce's Literary Texts*

Jisu Ryu · Hye-Soo Lee · Hyenyeong Chung · Moongee Jeon (Konkuk University)



This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: June 28, 2026
Revised: August 8, 2026
Accepted: August 23, 2026

Ryu, Jisu (First author)
Lecturer, Department of English
Language & Literature
Konkuk University
E-mail: jsryu0508@konkuk.ac.kr

Lee, Hye-Soo
Professor, Department of English
Language & Literature
Konkuk University
E-mail: hyesoo@konkuk.ac.kr

Chung, Hyenyeong
Researcher, The Konkuk Research
Institute for Multilingualism and
Multiculturalism
Konkuk University
E-mail: chn4050@konkuk.ac.kr

Jeon, Moongee (Corresponding author)
Professor, Department of English
Language & Literature
Konkuk University
120 Neungdong-ro, Gwangjin-gu,
Seoul, Korea
Tel: +82-2-450-3356
E-mail: mjeon1@konkuk.ac.kr

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2023S1A5B5A17085507).

ABSTRACT

Ryu, Jisu, Hye-Soo Lee, Hyenyeong Chung and Moongee Jeon. 2026. A corpus stylistic analysis of linguistic differences in James Joyce's literary texts. *Korean Journal of English Language and Linguistics* 26, 1117-1136.

This study conducts a corpus stylistic analysis of James Joyce's *Dubliners* and *A Portrait of the Artist as a Young Man* (hereafter *Portrait*). The two texts were selected for their literary significance and for exhibiting stylistic tendencies commonly associated with literary naturalism and modernism, respectively. While previous studies have discussed stylistic differences between these literary movements primarily through qualitative interpretation, relatively little research has attempted to examine such distinctions using quantitative linguistic evidence. Employing 26 indices from Coh-Metrix, a multilevel corpus analysis tool, this study investigates how linguistic features reflect the contrasting narrative strategies and thematic concerns of the two texts. The results reveal statistically significant differences across most linguistic measures. *Portrait* exhibits greater syntactic complexity, lower readability, and lower cohesion, reflecting the introspective and fragmented nature of modernist narrative and its use of stream-of-consciousness techniques. In contrast, *Dubliners* demonstrates greater discourse cohesion, lexical accessibility, and environmental grounding, consistent with stylistic tendencies associated with literary naturalism. The text also shows more frequent lexical repetition and negation, features that reinforce its thematic concern with paralysis and determinism. Additional differences in pronoun use, temporal cohesion, and spatial reference further distinguish the two works at both sentence and discourse levels. These findings provide empirical support for previously proposed stylistic distinctions between naturalist and modernist writing and demonstrate the value of quantitative corpus stylistic methods as a complement to traditional qualitative approaches in literary analysis.

KEYWORDS

literary texts, corpus stylistics, multilevel text analysis, Coh-Metrix

1. Introduction

A recent understanding of style defines it as a characteristic of texts made up of a combination of formal features that can be observed both quantitatively and qualitatively (Herrmann et al. 2015). Style has been studied across multiple disciplines, including corpus stylistics using computer tools, literary studies, and stylometry. Given the crucial role of style in literary texts and the increasing use of digital tools in literary research, many scholars in recent years have investigated the style of literary works by applying computational linguistic methods (Fischer-Starcke 2009). This approach to literature had been previously challenged by many skeptics in literary circles due to its apparent reductionism (Can and Patton 2004, Graesser et al. 2011). Many studies, however, have followed that understanding linguistic patterns by the corpus stylistic analysis is central to the illumination of literary meaning; language reflects main thematic ideas, subject matters, and characterization of a literary work, and furthermore manifests its genre (e.g., tragedy, comedy, and history) and discourse type (e.g., dialog vs. monologue). The idea is that a text addressing a particular subject matter, for example, has a thematic distinctiveness, which in turn provides evidence for linguistic idiosyncrasies. Thus, examining linguistic patterns is expected to reveal thematic elements of the text (Hardie and van Dorst 2020).

Corpus stylistic approaches are purported to offer several methodological benefits over traditional non-computer-mediated approaches. First, corpus stylistic analyses help identify otherwise undetected linguistic features, thus enriching literary interpretation (Murphy 2015). Second, corpus stylistic approaches afford an objective framework to empirically evaluate claims made by literary scholars (Stubbs 2005). Given these benefits, an increasing number of literary studies have adopted corpus stylistic approaches for literary interpretation over recent decades (e.g., Johnson 2016, Mahlberg 2013).

Previous studies in corpus stylistics, however, were exclusively concerned with concordance-driven analysis of literary texts (Fischer-Starcke 2009, Mahlberg 2013). For instance, Fischer-Starcke (2009) examined keywords and lexical bundles in *Pride and Prejudice*, and demonstrated how the central themes of the text can be uncovered by objective linguistic patterns. Notwithstanding the unique contribution of the previous studies to literary corpus stylistics, they were mainly biased toward lexical analyses. Lexical-based analyses, however, do not reveal the entire aspects of rhetorical choices made in a given text; the rhetorical choices of a work and their effects on writing styles and readers can be better understood with a comprehensive set of linguistic features including deeper-level ones.

Researchers have begun to take an interest in conspicuous advances in corpus linguistics, computational linguistics (Jurafsky and Martin 2008), and discourse processing (Kintsch 1988). The advancement in these fields combines to potentially enable stylists to analyze literature on a myriad of linguistic features, thereby relating linguistic patterns to the aesthetic dimensions of literary works (Graesser et al. 2011). More importantly, newly available analytical tools (e.g., Coh-Metrix) can be applied to analyze cohesion-related features in a text which can affect readers' processing of the text; note that it is one of the main goals of literary corpus stylistics to address the potential effect of language use on the text processing of readers (Mahlberg 2013).

The present study employs a corpus stylistic analysis to assess the early texts of James Joyce, *Dubliners* (Joyce 2008b) and *Portrait* (Joyce 2008a) (For convenience, subsequent references to Joyce (2008a, 2008b) are omitted after the first citation). It examines the linguistic features at the word, sentence, and discourse levels through a corpus analysis software, Coh-Metrix which has emerged as one of the most widely utilized tools in corpus and computational linguistic analysis (Graesser et al. 2004), extending beyond the traditional focus on word- and sentence-level analysis (Mahlberg 2013).

The comparison of the two texts is particularly intriguing given their distinct characteristics. James Joyce,

widely recognized as a leading modernist novelist of the twentieth century, fully embraces modernism's techniques in *Portrait*, including the "stream of consciousness." Conversely, his earlier work, *Dubliners*, shows many traits typical of naturalism, a movement to which numerous critics attribute *Dubliners* as a quintessential example (Holman and Harmon 1986).

This study postulates that two literary movements significantly impact the distinctive linguistic patterns observed in Joyce's texts. The current study strives to linguistically analyze the stylistic differences between *Dubliners* and *Portrait* using a variety of multilevel measures from Coh-Metrix. It aims to examine claims made by previous literary critics about naturalism and modernism, relating these movements' key characteristics to the linguistic features of the two texts.

2. Research Background

2.1 Previous Studies in Corpus-Stylistic Analysis of Literary Texts

Corpus stylistic methodology has been widely used in examining a variety of literary as well as nonliterary texts; for example, in identifying the authorship by distinctive language use of individual authors (Graesser et al. 2011, Pennebaker and King 1999), distinguishing genre-specific language (Lee and Jeon 2019) and discipline-specific language (Graesser et al. 2011), demonstrating literary characterization through key words and lexical clusters (Mahlberg 2013), describing characteristic language patterns in soliloquies versus dialogues in Shakespearean dramas (Murphy 2015), and identifying recurring themes in part I and II by comparing one to the other in a literary work (Johnson 2016) and linguistic idiosyncrasies in contemporary present-tense narratives versus past-tense narratives from the 20th century (Ikeo 2019).

The conventional techniques in corpus stylistic studies include analyses for concordance-based keywords, lexical bundles, and collocations (Mahlberg 2013, Stubbs 2005). The general procedure is to extract words more frequently used in a target corpus relative to a reference corpus. In addition to individual keywords, frequent lexical bundles and collocations (i.e., words that frequently co-occur together) are also subject to analysis (Mahlberg 2013). They are then grouped into certain semantic categories, which are critically interpreted with the reference to its literary meaning (Stubbs 2005). In the field of corpus stylistics, traditional methodologies focus on analyzing concordances to extract keywords, lexical bundles, and collocations (Murphy 2015). This involves identifying linguistic expressions that appear more frequently in a target corpus than in a reference corpus. The identified keywords, bundles, and collocations are then examined manually to assess their literary significance within the broader text context.

In Fischer-Starcke's study, for example, Jane Austen's *Pride and Prejudice* is analyzed through a critical examination of its keywords and lexical bundles, which ultimately illuminates the text's literary significance. The keyword analysis reveals that the novel's main themes, as indicated by the keywords, focus on characters, family relationships, and communication, with a particular emphasis on gender portrayals. These findings suggest that the novel primarily explores the complexity of human relationships. In the collocation analysis, which is designed to identify evaluative associations, the usage of words related to money and marriage suggests that family members are often linked to social connections (Fischer-Starcke 2009). This observation reflects the central thematic concerns of *Pride and Prejudice*, emphasizing the functional role of daughters in expanding social networks.

Murphy (2015) also employed a corpus stylistic approach to analyze soliloquies across 37 Shakespearean dramas, revealing distinct linguistic characteristics that differentiate soliloquies from dialogue. Key findings

included language pertaining to mental states and physicality, frequent use of exclamations such as 'O', and frequent use of first-person pronouns. These elements suggest that soliloquies serve primarily to express introspection and intense emotions. These studies have made unique contributions to the field of literature by uncovering implicit meanings from linguistic patterns in literary works. Their analysis, however, is largely based on superficial elements of language; in other words, they limited the scope of analysis to individual words (and a combination thereof) directly derivable from the text.

Researchers have pointed out that the traditional analytical approaches, which are lexically biased, can be limited in revealing various dimensions of language use (Graesser and McNamara 2011, Graesser et al. 2011). Recent attempts have been made to look beyond individual words and extract quantifiable linguistic features out of syntactically annotated corpora (Hardie and van Dorst 2020, Ikeo 2019). Hardie and van Dorst (2020), for example, examined 15 quantifiable syntactic features (e.g., pronouns, infinitives, and that-clause verb complements) in two syntactically tagged corpora, one of Shakespearean plays and the other by playwrights of the same period. The relative frequency of the features indicated that unique grammatical patterns were observed across plays, genre (e.g., comedy, history, and tragedy), and Shakespeare versus other playwrights; for example, Shakespearean dramas were characterized by a dominant use of tense, aspect, and pronoun features which indicated narrativity, relative to other playwrights. This study expands the existing scope of analysis in the corpus stylistic research by adding the syntactic features to the previous lexical-based analyses. Nevertheless, these studies still fail to address deeper-level features of language (e.g., cohesion). It has been emphasized that for a comprehensive literary analysis, corpus stylists need to broaden the current, lexically biased analytical framework to embrace more diverse features of language including cohesion-related measures as well as diverse lexical and syntactic indices (Graesser et al. 2011).

A paucity of studies using diverse measures is purported to be due to the methodological limitations of prevalent analytical software (e.g., WordSmith Tools). With significant advances in several fields of linguistics (e.g., corpus linguistics, computational linguistics, and discourse processing) emerged more sophisticated analysis tools. Coh-Metrix, one of the most widely used in corpus and computational linguistic analyses, for example, provides over 100 indices of linguistic and psycholinguistic features including traditional indices (e.g., word length, sentence length, and lexical diversity) and deeper-level indices (e.g., cohesion). More importantly, the tool has been validated on its predictability regarding the effects of textual features on readers' behavioral responses (reading and processing time), comprehension, memory, and perception of the text, both literary and nonliterary (Duran et al. 2007, Graesser et al. 2011). Duran et al. (2007) had experts rate 150 narrative and expository texts on their perceived temporal coherence. They found that the experts' ratings were significantly correlated with the automated analysis output on the temporal cohesion by Coh-Metrix. These studies suggest that it is possible to gauge the stylistic effects on readers' perception of a literary text by cohesion and other deep-level textual features inherent in the work.

As such advanced analytical tools become available, researchers have begun to use a large number of linguistic and psycholinguistic features in various types of literary stylistic analysis (e.g., Graesser and McNamara 2011). The majority of these studies, however, had a primary focus on the identification *per se* of the distinct linguistic features across different literary texts or genres; as a consequence, the identified linguistic patterns were not sufficiently complemented by literary interpretation. However, as emphasized in literary corpus stylistics (Fischer-Starcke 2009, Mahlberg 2013), it is far more revealing to further associate identified textual features with critical literary interpretation. The present study adopts a corpus stylistic approach to examine James Joyce's early literary works, *Dubliners* and *Portrait*. Specifically, it employs Coh-Metrix to investigate linguistic features at the lexical, syntactic, and discourse levels (Graesser et al. 2004).

2.2 Distinctive Characteristics of Naturalism and Modernism

Previous studies in corpus stylistics have shown that diverse linguistic features vary according to authors, genres, or themes of a literary text. These distinct linguistic features serve as a major stylistic device that reflects the major themes of literature. Accordingly, it stands to reason that a literary text of a particular literary movement would also be delineated by distinct linguistic features which in turn manifest the central motifs of the literary movement (Sotirova 2013). Literary texts in a given literary movement share similar philosophical, psychological, thematic, and aesthetic elements, which affect the idiosyncratic language use of the texts.

A growing number of literary scholars have begun to recognize the distinctiveness in linguistic patterns commonly found in a particular literary movement (Shiach 2007, Simion 2013, Sotirova 2013). However, few have attempted to empirically validate the linguistic distinctiveness of each literary movement with objective evidence to date. Therefore, it is the goal of the current study to substantiate previous literary discussion on literary movements by analyzing literary texts that belong to distinctive literary movements, drawing upon a comprehensive set of linguistic features. To this end, this study chooses two texts by James Joyce, one of the leading modernist writers, *Dubliners* and *Portrait*.

The reason that Joyce's texts are selected is as follows. First, as researchers suggest that it is of particular literary significance to conduct linguistic analyses on modernist texts due to their innovative language use (Mukhitdinova 2019, Simion 2013, Sotirova 2013), it is pivotal to examine the texts of Joyce, one of the most representative novelists of European modernism. Also, they further argue that its stylistic innovation is well captured by contrasting it with its literary predecessor (i.e., realism). Note that naturalism is a subfield of realism and therefore inherited several characteristics. Second, Joyce's two novels of *Dubliners* and *Portrait* show the characteristics of naturalism and modernism respectively. While modernism and naturalism are usually thought to be two distinctive literary movements, as far as the distance of Virginia Woolf from Emile Zola, Joyce's early novels interestingly displays the distinctive traits of each literary movement. Lastly, given that *Dubliners* and *Portrait* were written by the same author and published only two years apart, it is possible to observe the stylistic differences that are primarily induced by disparate literary movements with the other potential variables controlled for (i.e., author or time). The following presents a brief discussion of how each text represents characteristics of two literary movements along with their distinctive stylistic features.

The two literary movements relevant to the present study, literary naturalism in late nineteenth-century Europe and modernism in the twentieth century, display stark contrasts in their linguistic styles and worldviews. As Abrams cogently outlines in his classical *Glossary of Literary Terms*, naturalism depicts a human being whose "character and behavior are entirely determined by two forces, heredity and environment," and who is typically fated to a tragic end after being a pawn to various compulsions and an overwhelming environment (Abrams 1993). This pessimistic determinism underpins the distinctive features of naturalist texts, including themes of survival, paralysis, and a dry style. This naturalist movement is exemplified in Thomas Hardy's *Jude the Obscure*, which precedes Joyce's *Dubliners*. Compared to modernism, naturalism incorporates some features of realism, such as a realistic depiction of everyday life, a linear chronological narrative, verisimilar character descriptions, and detailed settings (Thornborrow and Wareing 1998). Specifically, one of the characteristics of naturalism in literature lies in its focus on the cold and objective description of the irresistible and overwhelming force of the external surroundings over the lives of the characters; it flourishes the deterministic and pessimistic view. In *Dubliners*, Dublin features as the oppressive setting that determines the characters' helpless and dismal fates. Dublin is the space of the famous theme of "paralysis" of *Dubliners*, which Joyce himself explains in his letter as the main focus of his novel (Kim 2003). These thematic ideas are stylistically expressed in a plain, not ornate, language, such as

everyday words or reiterating structures, and also in a monotonous and repetitive tone. In Joyce's own term, *Dubliners* is written in a style of scrupulous meanness. The simplicity and flatness of tone throughout *Dubliners* is Joyce's deliberate rhetorical style to best convey the themes and subject matter of the text. His rhetorical choices are, in turn, likely to contribute to shaping readers' perception of the surroundings and the characters therein (Wales 1992).

In contrast, modernist literature reflects the somber moods following World War I and almost obsessively pursues the value of novelty, adopting the "stream of consciousness," an uninterrupted flux of feelings and thoughts, as its dominant style and literary technique. It represents a shift in the perception of reality, not as an objective external world, but as the internal and flowing (hence "stream") consciousness of an individual. Since it serves as a key narrative method in modernist fiction, such as Joyce's *Portrait*, the stream of consciousness strives to depict, without the narrator's intervention, the full range and ongoing flow of a character's mental processes (Abrams 1993). Modernist texts are often complex and challenging, depicting the unrestricted flow of thoughts via the stream of consciousness (Sotirova 2013). Specifically, *Portrait* is a Künstlerroman (i.e., Bildungsroman of an "artist") that depicts the moral, religious, and intellectual awakening of a young artist, Stephen Dedalus, who rebels against the oppressive religious and national conventions. It was published in a book form in 1916. Several modernist elements are found in *Portrait*, such as the rebellion against the established order, depressing conventions, and oppressive faith (Simion 2013) or a concern with the individual's psychic reality rather than external situations (Sotirova 2013). *Portrait* also demonstrates several important modernist experiments of style and form, particularly including the famous 'stream of consciousness.' As many scholars have noted, stream of consciousness is one of the most defining techniques of modernist novels (Sotirova 2013) which records multitudinous thoughts and emotions of the characters, and thus presents the narrative world through the viewpoint of characters' consciousness rather than objective reality (Simion 2013). One of the stylistic consequences of modernist techniques in *Portrait* is its linguistic intricacy; critics argue that Joyce adopts ornate language in this modernist text (e.g., complex structures and vocabulary) which matches thematic intricacies of it (i.e., moral chaos before awakening) (Sotirova 2013). Stylistic effects of stream of consciousness induce low cohesion (particularly temporal cohesion). Note in passing that cohesion refers to objective text features indicating connectivity among the elements of the text; cohesion, in turn, affects text coherence which is readers' perception of the textual connectivity and smoothness (Duran et al. 2007). Thus, these stylistic choices contribute to shaping readers' perception of the textual cohesion. Although much attention has been paid to *Dubliners* and *Portrait*, few systematic attempts have been made with empirical data to substantiate how two texts reflect the differences of two different literary movements of naturalism and modernism. It is at this point that this study intervenes with the methodology of corpus stylistics.

Despite the close relationship between the linguistic features of literary texts and their corresponding movements, few scholars have undertaken empirical studies to determine how different styles are linguistically represented by different literary movements. However, exceptions such as the research conducted by Totò (2014) and Stubbs (2005) exist. For example, Totò examined the linguistic features of the stream-of-consciousness technique in modernist texts through manual analysis. He identified several linguistic traits of the technique, such as the absence of punctuation, unconventional use of connectors, and the ellipsis of pronouns and prepositions, which contribute to the impression of free-flowing random thoughts. These earlier studies, including his investigation, indicate that the principal themes and techniques commonly used in different literary movements can be correlated with specific linguistic features.

3. Methodology

3.1 Literary Corpora for the Current Study

The present study conducts a comparative analysis of James Joyce's *Dubliners* and *Portrait*, focusing on various linguistic features. The original novels, accessed online through the Gutenberg Project website (<http://www.gutenberg.org/>), form the basis of this research. The Gutenberg Project offers a wide range of out-of-copyright electronic literary texts to the public. The corpus construction procedure involves saving each chapter as a text file (i.e., txt files). *Dubliners* comprises fifteen separate chapters with distinct short stories, whereas *Portrait* includes five continuous chapters. The next step involves removing extraneous parts of the text that do not pertain to the analysis, such as chapter titles and special characters (e.g., asterisks). Subsequently, the text files of each chapter are split into smaller files for analysis. Considering the effects of text length and the file size limits of Coh-Metrix, this study standardized each file to approximately 400 words. Table 1 presents the dimensions of the two corpora. As indicated in Table 1, the *Dubliners* corpus comprises 156 files (62,350 words) and the *Portrait* corpus comprises 196 files (81,182 words). Although the sample sizes for the two conditions in this study are unequal, the comparison remains statistically valid. This validity is due to the assumption of homogeneity of variance, essential for the analysis of variance (ANOVA) used in this study, which remains unaffected by the unequal sample sizes (Kim and Cribbie 2018). Although this study conducts statistical analyses to compare the two works, it would like to note in advance that, since it analyzes only two types of texts, the examination of linguistic differences between Naturalism and Modernism is conducted at an exploratory level.

Table 1. Size of *Dubliners* and *Portrait* Corpora

Measures	<i>Dubliners</i>	<i>Portrait</i>
Total Number of Files	156	196
Total Number of Words	62,350	81,182
Mean Number of Words	400	414
Mean Number of Sentences	28.7	27.2

3.2 A Brief Sketch of *Dubliners* and *Portrait*

Dubliners (1914) is a collection of fifteen interconnected short stories that explore the lives of Dublin's residents. Regarded as a quintessential naturalist text in its thematic and stylistic elements, *Dubliners* aligns with the naturalism portrayed (Kim 2003, Lehan 1994) in literary works like Emile Zola's *Nana*, which highlights a detached depiction of characters' lives heavily influenced by external factors and marked by a bleak sense of inevitability. In *Dubliners*, Dublin serves as an oppressive backdrop that profoundly shapes the characters' dismal and fated destinies. Joyce notably identified the poignant theme of 'paralysis' as central to *Dubliners* (Kim 2003), a theme expressed through the stark and straightforward language characteristic of naturalist literature.

Portrait (1916) is a *Künstlerroman* (i.e., *Bildungsroman* of an "artist"), which is a novel genre that focuses on the formative years of an artist. It traces the moral, religious, and intellectual development of the protagonist, Stephen Dedalus, as he endeavors to establish his identity amidst societal pressures and rigid religious doctrines. Serialized from 1914 to 1915 and released as a book in 1916, *Portrait* is celebrated as a seminal work of modernist literature (Lehan 1994). As a modernist piece, it challenges traditional norms and values, prioritizing the inner world (i.e., psychic reality) of the individual over external factors. One of its most notable experimental techniques is the "stream of consciousness" method, which captures the random flow of a character's thoughts and perceptions (Sotirova 2013).

3.3 Multidimensional Linguistic Indices and Research Hypothesis

The analytical tool employed in this study is the Coh-Metrix system, developed by the Institute for Intelligent Systems at the University of Memphis (Graesser et al. 2004). Coh-Metrix has been extensively utilized in fields like corpus stylistics, corpus linguistics, stylometry, and computational linguistics for analyzing various text types and discourses. In this study, 26 measures are used to examine the linguistic style features of the two texts at a multi-layered level (see Table 2). These measures have been previously employed in Coh-Metrix-based studies to explore different text and discourse features (Graesser and McNamara 2011). Notably, many studies have demonstrated that linguistic and psycholinguistic features measured via Coh-Metrix are highly correlated with human raters' perception of the text along with readers' memory and comprehension performances (Graesser and McNamara 2011). This suggests that it is possible to gauge the potential effects of diverse linguistic features on readers' perception of a literary text using this system. Due to these new possibilities in revealing features otherwise undetected, Coh-Metrix has been widely employed to analyze diverse types of text and discourse in corpus stylistics, corpus linguistics, stylometry, and computational linguistics.

Table 2. Coh-Metrix Indices Selected for the Current Study

Category	Subcategory	Coh-Metrix Index
Composite Textual Difficulty	Readability Indices	Flesch Reading Ease (FRE) Flesch-Kincaid Grade Level (FKGL)
Lexical Information	Word Frequency	Word Frequency (Content Words) Word Frequency (All Words)
	Lexical Diversity	Type-Token Ratio (Content Words)
	Word Features	Concreteness
		Age of Acquisition
		Familiarity
		Average Word Length
	Negation	Negations
Personal Pronouns		First-person Pronouns Second-person Pronouns Third-person Pronouns All Pronouns
Syntactic Complexity		Modifiers per NP Words Before Main Verbs Average Sentence Length
Cohesion	Reference Cohesion	Argument Overlap (Adjacent Sentences) Argument Overlap (All Sentences) Content Word Overlap (Adjacent Sentences) Content Word Overlap (All Sentences)
	Temporal Cohesion	Temporal Cohesion Aspect Repetition Score Tense Repetition Score
	Spatial Cohesion	Locational Entities Motion Prepositions

The current study suggests that these measures can be effectively used to study the linguistic features of two literary texts, *Dubliners* and *Portrait*, representing naturalism and modernism respectively. The effectiveness of some indices, influenced by text length, is estimated using a relative frequency score, which represents the occurrence of a specific category per 1,000 words (Graesser et al. 2004).

3.3.1 Composite textual difficulty

The analysis of composite textual difficulty includes readability indices. While traditional readability measures such as the Flesch Reading Ease (FRE) and Flesch-Kincaid Grade Level (FKGL) primarily focus on word and sentence length, their scores are inversely correlated. FKGL scores, ranging from 1 to 12, correspond to US grade levels. In contrast, FRE scores span from 0 to 100, with higher values indicating simpler texts. These measures have been widely utilized in corpus linguistics to gauge text difficulty (Graesser and McNamara 2011). These readability scores are calculated primarily on the basis of word and sentence length and provide a comprehensive measure of textual complexity that captures, albeit indirectly, aspects of vocabulary, syntax, and discourse-level organization. It is hypothesized that *Portrait* (a Modernist work) will exhibit higher FKGL scores and lower FRE scores than *Dubliners* (a Naturalist work).

3.3.2 Lexical information

Lexical analyses involve assessing word frequency for both content words and all words, as well as negations. The CELEX database, developed by the Dutch Centre for Lexical Information, provides a detailed compilation of word frequencies from over 17.9 million words (Graesser et al. 2004). Here, the raw frequency score undergoes logarithmic transformation, which correlates more closely with reading times and offers a more precise evaluation of word difficulty compared to the raw frequency score (Just and Carpenter 1987). The negation index measures the relative frequency of negation expressions.

Lexical information includes dimensions such as the lexical diversity index, measured by the type-token ratio (TTR) for content words. A “type” signifies the unique words in a text, whereas a “token” denotes the instances of the type. A text with more repeated words typically shows a decrease in TTR, reflecting lower lexical diversity. An additional dimension concerns word features, which in Coh-Metrix are assessed using the MRC Psycholinguistic Database (Coltheart 1981), spanning ratings from 100 to 700. The MRC database offers ratings of several thousand words by adults on various psychological dimensions (Graesser et al. 2004). Words are rated on a 7-point scale for concreteness, familiarity, and age of acquisition (AOA); scores are then multiplied by 100 to obtain the final score. Higher scores for concreteness and familiarity make a word easier to comprehend, while a lower AOA score indicates easier word processing. In addition, lexical information measures include average word length, with longer words generally indicating greater lexical complexity and, consequently, greater textual difficulty (Graesser and McNamara 2011).

Dubliners and *Portrait*, influenced by distinct literary movements, exhibit unique lexical styles. *Dubliners* employs simple and concrete language that realistically portrays the lives of working-class characters, utilizing familiar vocabulary for an objective depiction of everyday life (Thornborrow and Wareing 1998). Conversely, *Portrait* employs syntactically complex, extended, and stylistically experimental language that reflects the protagonist's evolving consciousness (Sotirova 2013). This study expects that less specific vocabulary and longer words will be used in *Portrait*. Additionally, themes of paralysis and repetitive futility in *Dubliners* may lead to reduced lexical diversity, increasing the repetition of words. Furthermore, the pervasive pessimism in *Dubliners* might be expressed linguistically through frequent negation.

3.3.3 Personal pronouns

The analysis of pronouns includes the relative frequencies of first-, second-, and third-person pronouns, along

with all personal pronouns (Hardie and van Dorst 2020). Personal pronouns play a crucial role in corpus stylistics, as they serve as key indicators of narrative perspective. By signaling the point of view from which a text is conveyed, they influence the mode of narration and contribute significantly to the way the text is interpreted by the reader (Bosseaux 2004). Given that *Dubliners* focuses on multiple characters, a higher frequency of pronouns is expected to characterize the diverse cast of characters. Conversely, *Portrait* primarily employs the stream of consciousness technique, emphasizing the protagonist's internal thoughts and perceptions over interactions with others. Therefore, the use of personal pronouns in *Portrait* is likely to be less compared to *Dubliners* (Totò 2014).

3.3.4 Syntactic complexity

Coh-Metrix provides indices of syntactic complexity, which are assessed using two distinct measures: modifiers per NP (referring to the number of modifiers per noun phrase), also known as NP density, and words before main verbs (measuring the number of words before main verbs, also known as subject density). Higher scores on both indices signify greater syntactic complexity (Graesser et al. 2004). Measures of syntactic complexity also include average sentence length. As sentence length increases, syntactic complexity tends to increase (Graesser and McNamara 2011). As *Portrait* employs the stream-of-consciousness technique (Sotirova 2013), it is anticipated to exhibit higher syntactic complexity than *Dubliners*.

3.3.5 Cohesion

Coh-Metrix provides several cohesion-related features, including reference, temporal, and spatial cohesion. Cohesion is defined as the extent to which linguistic elements within a text are explicitly connected (Halliday and Hasan 1976), implying that a text with high cohesion exhibits more robust connections between its linguistic units than a text with low cohesion.

Reference cohesion calculated by argument overlap occurs when the same noun (in singular or plural form) or pronoun appears in adjacent sentences or across all possible sentence pairs (Graesser et al. 2004). Another measure of reference cohesion, the content word overlap, indicates the proportion of content words that recur between adjacent sentences or in all sentence pairs.

The temporal cohesion measure refers to the average repetition score for tense (e.g., *past*, *present*, *future*) and the aspect repetition score (e.g., *perfective*, *progressive*). The tense repetition score is a ratio calculated by dividing the number of adjacent sentences with the same tense by the total number of sentence pairs. The aspect repetition ratio is determined by dividing the number of adjacent sentences with the same aspect by the total number of sentence pairs. Temporal cohesion, aspect repetition, and tense repetition are closely related measures in that they are all derived from the distribution and repetition of verbal tense and aspect, but they differ in the level at which repetition is assessed. Tense repetition measures the extent to which the same grammatical tense is repeatedly used across successive verbs in a text (e.g., repeated use of the past tense), whereas aspect repetition measures the extent to which the same aspectual category is repeated (e.g., completed, in-progress, or static). In contrast, temporal cohesion provides a composite measure that captures the overall repetition of temporal information by combining these two dimensions. Specifically, as indicated by the definition of temporal cohesion, it is calculated by averaging the repetition score for tense and the repetition score for aspect. Thus, while tense repetition and aspect repetition represent separate components of temporal organization, temporal cohesion integrates both to provide an overall measure of the consistency or continuity of temporal relations among events and actions across the text.

The spatial cohesion of Coh-Metrix includes two indices: location entities (e.g., *place*, *region*, *beside*, *here*,

there) and motion prepositions (e.g., into, through) in WordNet, both calculated based on the relative frequency of a category per 1,000 words. WordNet provides classification information for motion and location words (Graesser et al. 2004).

It is expected that *Portrait* will exhibit lower levels of referential, temporal, and spatial cohesion compared to *Dubliners*. This is because *Portrait* employs the stream of consciousness technique, which produces random associations between sentences, resulting in lower levels of argument and content word overlap (Sotirova 2013). Additionally, this technique disrupts the temporal progression of the narrative, likely resulting in fewer tense and aspect repetitions compared to *Dubliners*. Ultimately, modernist texts like *Portrait* prioritize internal thought associations, while naturalist texts like *Dubliners* focus on detailed representations of the external environment (Thornborrow and Wareing 1998). Therefore, it is hypothesized that the spatial cohesion in *Dubliners* will surpass that in *Portrait*.

3.4 Statistical Analyses

The aim of this study is to examine the stylistic differences between James Joyce's *Dubliners* and *Portrait* by comparing 26 Coh-Metrix measures. Inferential statistics, utilizing one-way analysis of variance (ANOVA), are employed to test the previously formulated hypotheses (Kim and Cribbie 2018). Here, the corpus type, specifically *Dubliners* and *Portrait*, is the independent variable, with the 26 Coh-Metrix measures serving as dependent variables. Hypothesis testing is conducted at a 5% significance level, adhering to Fisher's convention (Rao 1992).

4. Results and Discussion

4.1 Composite Textual Difficulty

Table 3 presents the results of the statistical analysis of composite textual difficulty in the *Dubliners* and *Portrait* corpora. The analysis revealed that *Portrait* exhibits higher Flesch–Kincaid Grade Level (FKGL) scores and lower Flesch Reading Ease (FRE) scores than *Dubliners*. These findings suggest that the greater linguistic complexity of *Portrait*, as reflected in its longer words and sentences, may be associated with Joyce's use of the stream-of-consciousness technique (Sotirova 2013). This is evident in example (1), a characteristic of modernist texts (Sotirova 2013). Example (1) from *Portrait* captures moments of epiphany, which Joyce described as a 'sudden spiritual manifestation ... in a memorable phase of the mind itself' (Joyce 2008a, p. 211). In this passage, the protagonist Stephen Dedalus undergoes a stream of consciousness, realizing the symbolic meaning of his name. Specifically, Stephen's name symbolizes Daedalus from Greek mythology, who crafted wings for himself and his son, Icarus. Thus, his name gains deeper significance, contributing to the overarching themes of the novel, *Portrait*.

- (1) What did it mean? Was it a quaint device opening a page of some medieval book of prophecies and symbols, a hawklike man flying sunward above the sea, a prophecy of the end he had been born to serve and had been following through the mists of childhood and boyhood, a symbol of the artist forging anew in his workshop out of the sluggish matter of the earth a new soaring impalpable imperishable being? (Ch. 4 of *Portrait*)

In contrast, *Dubliners* is marked by a clear and succinct style (Thornborrow and Wareing 1998), as evidenced

by readability indices. This stylistic approach is consistent with the core characteristics of naturalist literature, which often portrays a pessimistic and deterministic worldview, depicting individuals as subject to the inescapable forces of their environment and nature (Kim 2003). The following example (2) from *Eveline* demonstrates that Eveline is caught in a state of paralysis, unable to leave Ireland even when urged by her lover, Frank, to board a ship for South America. Instances of paralysis are central to the overall theme of the text and are emblematic of the key elements of naturalist literature (Kim 2003). They are expressed in a detached and succinct style, as illustrated by Eveline's depiction as 'passive, like a helpless animal.'

- (2) He was shouted at to go on but he still called to her. She set her white face to him **passive, like a helpless animal**. Her eyes gave him no sign of love or farewell or recognition ("Eveline" of *Dubliners*)

Table 3. Analysis of the Readability Index in *Dubliners* and *Portrait* Corpora

Index	<i>Dubliners</i> (1914) (<i>n</i> = 156)	<i>Portrait</i> (1916) (<i>n</i> = 196)	<i>F</i>	<i>p</i>
Flesch Reading Ease	78.2 (8.35)	74.1 (12.3)	12.865	.000*
Flesch-Kincaid Grade Level	6.1 (2.2)	7.1 (2.9)	13.074	.000*

Note. The figures in parenthesis indicate the standard deviation. **p* < .05

4.2 Lexical Information

Table 4 presents the results of the lexical analysis conducted on the *Dubliners* and *Portrait* corpora. The frequency of both content words and total words is significantly higher in *Dubliners* compared to *Portrait* (*ps* < .05). Moreover, the type-token ratio (TTR) of *Dubliners* is significantly lower than that of *Portrait* (*p* < .05), suggesting a greater degree of lexical repetition in *Dubliners*.

Although no statistically significant differences were observed between the two corpora in terms of concreteness or average word length (*ps* > .05), *Dubliners* exhibited higher levels of word familiarity and a lower mean age of acquisition. These patterns may be attributed to Joyce's frequent use of familiar, commonly used vocabulary in *Dubliners*, which contributes to its lower TTR. Furthermore, the lower age of acquisition score aligns with the vernacular speech of the middle- and working-class characters portrayed in the text (Clee 1965).

Table 4. Analysis of the Lexical Information in *Dubliners* and *Portrait* Corpora

Index	<i>Dubliners</i> (1914) (<i>n</i> = 156)	<i>Portrait</i> (1916) (<i>n</i> = 196)	<i>F</i>	<i>p</i>
<i>Word Frequency</i>				
Word Frequency (Content Words)	2.36 (0.20)	2.20 (0.20)	56.032	.000*
Word Frequency (All Words)	3.17 (0.08)	3.13 (0.09)	20.778	.000*
<i>Lexical Diversity</i>				
Type-Token Ratio (Content Words)	0.76 (.08)	0.78 (0.08)	4.883	.028*
<i>Word Features</i>				
Concreteness	397 (26.9)	398 (28.2)	0.217	.642
Age of Acquisition	302 (29)	313 (30)	11.843	.001*
Familiarity	576 (6.60)	567 (9.52)	99.399	.000*
Average Word Length	1.34 (0.06)	1.35 (0.07)	2.633	.106
<i>Negation</i>				
Negations	10.9 (0.64)	9.50 (0.57)	2.631	.106

Note. The figures in parenthesis indicate the standard deviation. **p* < .05

Specifically, example (3) clearly demonstrates how the character's use of simple language is linked to their state

of paralysis (Kim 2003).

- (3) a. ...though Joe's wife was ever so **nice** with her...
 b. She used to have such a bad opinion of Protestants but now she thought they were very **nice** people, a little quiet and serious, but still very **nice** people to live with
 c. She hoped they would have a **nice** evening
 d. [Joe] was very **nice** with her ("Clay" of *Dubliners*)

Maria, the protagonist of "Clay," is a poor, elderly, and unmarried woman struggling to effectively express her feelings. Frequently using the word 'nice', she describes various situations including Joe's wife, her co-workers, the weather, and even Joe, who remains oblivious to her emotions. Maria's simple linguistic style reflects her passive engagement with her surroundings and her submission to external factors, accentuating the theme of paralysis, a central element in both naturalist literature and *Dubliners* overall (Thornborrow and Wareing 1998).

The comparison of negation-related measures indicates no statistically significant differences between the two corpora ($p > .05$). However, descriptive statistics show a slightly higher frequency of negation in *Dubliners*, which may be interpreted as reflective of the text's generally pessimistic tone—a hallmark of naturalist literature (Gibbons 2000).

4.3 Personal Pronouns

Table 5 presents the results of the statistical analysis of personal pronoun usage in the *Dubliners* and *Portrait* corpora, revealing significant differences across all pronoun categories. With the exception of second-person pronouns, which are infrequently used in both texts, *Dubliners* shows significantly higher frequencies of first-person, third-person, and overall personal pronoun usage ($ps < .05$). This pattern aligns with the text's naturalistic characteristic, which emphasizes realistic and detailed portrayals of a broad range of characters in *Dubliners*. In contrast, *Portrait* centers predominantly on the internal perspective of its protagonist, Stephen Dedalus, with minimal narrative attention given to other characters (Sotirova 2013).

While *Dubliners* frequently uses pronouns to describe characters particularly in scenes with multiple characters, proper names are sometimes employed to avoid ambiguity, as in '...ever since Kate and Julia, following their brother Pat's death, had moved from the house in Stoney Batter to live with Mary Jane...' ("The Dead" in *Dubliners*). Nevertheless, pronouns are more commonly used to depict characters in *Dubliners*, as shown in examples (2), (3), and (6). Conversely, examples (1), (4), and (7) from *Portrait* emphasize the central role of Stephen Dedalus, thus reducing the need for pronouns in character descriptions.

Table 5. Analysis of the Personal Pronouns in *Dubliners* and *Portrait* Corpora

Index	<i>Dubliners</i> (1914) ($n = 156$)	<i>Portrait</i> (1916) ($n = 196$)	F	p
First-person Pronouns	10.3 (12.5)	7.1 (8.2)	8.315	.004*
Second-person Pronouns	3.82 (4.54)	5.34 (6.99)	5.525	.019*
Third-person Pronouns	40.7 (16.1)	34.1 (13.9)	17.137	.000*
All Pronouns	54.8 (15.8)	46.5 (14.5)	26.239	.000*

Note. The figures in parenthesis indicate the standard deviation. * $p < .05$

It is noteworthy that both corpora show a consistently low frequency of second-person pronoun usage. In *Dubliners*, the narrative perspective shifts between first- and third-person narration, with the first three stories

employing a first-person narrator and the later stories adopting a third-person point of view. Similarly, *Portrait* predominantly employs a third-person narrative perspective, with occasional shifts to the first-person in the later chapters. These narrative strategies reflect Joyce's stylistic emphasis on objective narrative distance.

4.4 Syntactic Complexity

Table 6 presents the results of the syntactic analysis conducted on the *Dubliners* and *Portrait* corpora, revealing statistically significant differences across three measures of syntactic complexity ($ps < .05$). Specifically, *Portrait* demonstrates greater syntactic complexity than *Dubliners*, as indicated by its higher noun density, subject density, and average sentence length.

Table 6. Analysis of the Syntactic Complexity in *Dubliners* and *Portrait* Corpora

Index	<i>Dubliners</i> (1914) ($n = 156$)	<i>Portrait</i> (1916) ($n = 196$)	F	p
Modifiers per NP	0.76 (0.15)	0.80 (0.15)	6.578	.011*
Words Before Main Verbs	2.84 (1.13)	3.40 (1.91)	10.345	.001*
Average Sentence Length	15.2 (4.76)	18.3 (8.53)	16.425	.000*

Note. The figures in parenthesis indicate the standard deviation. * $p < .05$

From a literary perspective, *Portrait* is emblematic of modernist literature, which is often characterized by stylistic and structural complexity—particularly through the use of the stream-of-consciousness technique (Sotirova 2013). The earlier example (1) demonstrates this through its depiction of numerous thoughts and ideas lacking clear causality. The simultaneous presence of numerous different thoughts results in the use of complex noun phrases such as “a prophecy of the end he had been born to serve and had followed through the mists of childhood and boyhood” and “a symbol of the artist forging anew in his workshop from the sluggish matter of the earth a new, soaring, impalpable, imperishable being.” In another example (4), Stephen discusses his theory of aesthetics with his friends following an epiphany that reaffirmed his artistic vocation (cf. The subject of the sentence in this example is highlighted in bold to emphasize the numerous words before the main verbs).

(4) The instant wherein that supreme quality of beauty, the clear radiance of the esthetic image, is apprehended luminously by the mind which has been arrested by its wholeness and fascinated by its harmony is the luminous silent stasis of esthetic pleasure... (Ch. 5 of *Portrait*)

The examples (1) and (4) clearly demonstrate how the syntactically complex passages closely relate to the central technique of modernist literature, specifically, the stream of consciousness, and to the theme of artistic chaos. Previous research has consistently shown that modernist texts exhibit increased linguistic complexity, largely due to this narrative mode. The present study extends previous findings by confirming that the three selected measures of syntactic complexity effectively capture the intricate linguistic features associated with modernist texts.

4.5 Cohesion

Table 7 presents the results of the cohesion analysis conducted on the *Dubliners* and *Portrait* corpora. The analysis reveals statistically significant differences in argument and content word overlap scores across all sentence

pairs between the two corpora ($ps < .05$). In contrast, no significant differences were observed for adjacent sentence pairs ($ps > .05$). The results indicate that *Dubliners* exhibits a higher degree of reference cohesion than *Portrait*, as evidenced by the higher argument and content word overlap scores for all sentence pairs. This is in line with Wales's (1992) assertion that 'repetition' is a key stylistic device used by Joyce in *Dubliners*. She cites examples of Joyce employing lexical repetition to enhance themes and create a specific atmosphere, such as the repeated use of 'dark' to accentuate the gloomy and lifeless atmosphere in example (5).

- (5) The career of our play brought us through the **dark** muddy lanes behind the houses where we ran the gantlet of the rough tribes from the cottages, to the back doors of the **dark** dripping gardens where odours arose from the ashpits, to the **dark** odorous stables ... ("Araby" of *Dubliners* cited from Wales 1992)

Another example from *Dubliners*, specifically 'The Dead' (example 6), shows how the word 'death' is used three times in repetition in this, the longest and most famous finale of *Dubliners*.

- (6) a. ...ever since Kate and Julia, after the **death** of their brother Pat, had left the house in Stoney Batter and taken Mary Jane..."
 b. "I implored of him to go home at once and told him he would get his **death** in the rain. But he said he did not want to live."
 c. He did not like to say even to himself that her face was no longer beautiful but he knew that it was no longer the face for which Michael Furey had braved **death**. ("The Dead" in *Dubliners*)

These findings suggest that *Dubliners* exhibits a greater degree of referential cohesion compared to *Portrait*, as evidenced by higher levels of argument and content word overlap throughout the text. This observation supports Wales's claim that "repetition" functions as a key stylistic device in *Dubliners* (Wales 1992). It illustrates how Joyce strategically employs lexical repetition to reinforce thematic elements and to cultivate a distinctive narrative atmosphere.

All indices related to temporal cohesion, with the exception of aspect repetition, indicate significantly stronger temporal cohesion in *Dubliners* than in *Portrait* ($ps < .05$). The absence of significant difference in aspect repetition suggests that this particular feature is consistent across the two texts. The comparatively reduced temporal cohesion in *Portrait* is closely associated with the modernist hallmark of stream of consciousness narration (Sotirova 2013). In *Portrait*, Stephen's narrative frequently intertwines events from different time periods, mixing past, present, and future in a non-linear and often disruptive manner, thereby undermining conventional chronological progression. This blending of times is evident from a manual analysis of certain passages in *Portrait*, like the one in example (7).

- (7) "His soul **had arisen** from the grave of boyhood, spurning her graveclothes. Yes! Yes! Yes! He **would create** proudly out of the freedom and power of his soul, as the great artificer whose name he bore, a living thing, new and soaring and beautiful, impalpable, imperishable. He **started up** nervously from the stoneblock for he could no longer quench the flame in his blood" (Ch. 4 of *Portrait*)

In example (7), the stream of consciousness technique vividly depicts Stephen's deepest thoughts. The past perfect tense portrays Stephen as having emerged from childhood, and the use of the future tense "would" expresses his ambition to use his soul's power as a formidable artist in the future. This description unfolds as if

narrated by a third-person entity entering Stephen's consciousness. The narrative subsequently shifts to the past tense, depicting Stephen from a distant, third-person perspective.

Table 7. Analysis of Cohesion in *Dubliners* and *Portrait* Corpora

Index	<i>Dubliners</i> (1914) (<i>n</i> = 156)	<i>Portrait</i> (1916) (<i>n</i> = 196)	<i>F</i>	<i>p</i>
<i>Reference Cohesion</i>				
Argument Overlap (Adjacent)	0.45 (0.17)	0.44 (0.20)	0.225	.636
Argument Overlap (All)	0.38 (0.14)	0.34 (0.18)	3.991	.047*
Content Word Overlap (Adjacent)	0.11 (0.01)	0.10 (0.01)	1.652	.200
Content Word Overlap (All)	0.09 (0.01)	0.07 (0.01)	7.171	.008*
<i>Temporal Cohesion</i>				
Temporal Cohesion	0.87 (0.07)	0.85 (0.08)	5.966	.015*
Aspect Repetition Score	0.90 (0.11)	0.92 (0.09)	3.319	.069
Tense Repetition Score	0.85 (0.11)	0.79 (0.12)	22.060	.000*
<i>Spatial Cohesion</i>				
Locational Entities	3.33 (0.37)	2.08 (0.33)	6.348	.012*
Motional Prepositions	45.57 (13.03)	43.13 (12.00)	3.337	.069

Note. The figures in parenthesis indicate the standard deviation. * $p < .05$

In contrast, *Dubliners* maintains a more objective narrative stance, adhering to a linear temporal structure even when incorporating flashbacks, thereby reinforcing narrative clarity and coherence (Coleman 1991). This focus on chronological order reflects naturalist nature of the text, which is characterized by a deterministic and often pessimistic portrayal of external reality. The contrast between the fragmented temporality of modernist texts and the structured temporal progression typical of naturalist texts is effectively captured by the temporal cohesion indices used in this analysis.

Furthermore, the two corpora exhibit significant differences in spatial cohesion. The use of locational entities is significantly more frequent in *Dubliners* than in *Portrait* ($p < .05$), whereas the difference in the use of motional prepositions, although higher in *Dubliners*, does not reach statistical significance ($p > .05$). These findings suggest that *Dubliners* places greater emphasis on concrete geographical details, including specific references to architecture, street names, and urban spaces. This focus is consistent with a well-documented tendency in naturalist literature to foreground the physical environment as a determinant of human experience (Thornborrow and Wareing 1998). In contrast, *Portrait* shows a relative disengagement from external spatial description, prioritizing instead the internal, psychological landscape of its protagonist.

For instance, example (8) demonstrates how the detailed description of Dublin's geography is central to *Dubliners*.

- (8) **North Richmond Street**, being blind, was a quiet street except at the hour when **the Christian Brothers' School** set the boys free. **An uninhabited house of two storeys** stood at the blind end, detached from its neighbours **in a square ground**. **The other houses of the street**, conscious of decent lives within them, gazed at one another with brown imperturbable faces. ("Araby" in *Dubliners*)

In example (8), the locations are not mere backgrounds in the narrative; they are depicted as bleak, dreary, and underdeveloped, effectively illustrating the recurrent theme of naturalism: the characters' pervasive sense of paralysis and inability to act (Kim 2003). The focus on the spatial settings in *Dubliners* aligns with the text's naturalistic trend by highlighting the dominant influence of the environment on characters who are helpless to

change their circumstances.

5. Conclusion

This study provides empirical evidence, through the application of Coh-Metrix, that the stylistic difference between *Dubliners* and *Portrait* is closely linked to the distinct linguistic features associated with naturalism and modernism. The stylistic contrast between the two texts reflects the different priorities of their respective literary movements. *Dubliners* is characterized by a clear and succinct style, marked by detached and objective narration (Thornborrow and Wareing 1998), whereas *Portrait* adopts a more linguistically complex style, employing stream-of-consciousness techniques to convey the protagonist's inner experiences (Sotirova 2013). The stylistic complexity of *Portrait* is particularly salient at the sentence level, as evidenced by several quantitative indicators: longer average sentence lengths, higher Flesch-Kincaid Grade Level (FKGL) scores, lower Flesch Reading Ease (FRE) scores, a greater number of modifiers per noun phrase, and a higher number of words before the main verbs. These findings highlight the relationship between modernist narrative strategies and increased syntactic complexity.

Furthermore, the theme of paralysis in *Dubliners*, emblematic of its naturalist characteristic (Kim 2003), is linguistically expressed at both the lexical and discourse levels. At the lexical level, *Dubliners* shows a higher frequency of familiar words and a higher number of words acquired at an earlier age, as well as a lower type-token ratio (TTR) and more frequent use of negation. These features contribute to a linguistic style that reinforces the dark and deterministic tone of the text. At the discourse level, *Dubliners* exhibits greater cohesion, as reflected in higher argument and content word overlap scores (Wales 1992), further supporting the interpretation that linguistic repetition and referential cohesion are stylistically significant in the text.

The stylistic and linguistic contrasts between the two literary works also stem from their different approaches to characterization. *Dubliners* portrays a range of characters across multiple narratives, whereas *Portrait* is primarily concerned with the inner consciousness of a single protagonist. This divergence is linguistically reflected in the more frequent use of personal pronouns in *Dubliners*, indicative of its broader character focus.

In addition, the texts differ markedly in their temporal structuring. *Dubliners* generally adheres to a linear narrative framework (Leonard 1994), while *Portrait* adopts a non-linear, modernist approach that foregrounds the fragmented nature of memory, perception, and subjective experience (Sotirova 2013). These narrative strategies are reflected in linguistic measures of temporal cohesion, with *Dubliners* showing greater temporal clarity across multiple indices.

Finally, the two works contrast in their treatment of spatial detail. *Dubliners* places a greater emphasis on the external environment, as evidenced by significantly higher frequency of locational entities, a pattern consistent with naturalist literature's focus on setting (Thornborrow and Wareing 1998). In contrast, *Portrait* privileges the internal psychological landscape of its protagonist, reflecting the introspective characteristic of modernist narrative.

This study has several implications for the application of corpus stylistic methods to literary analysis. It represents an early empirical attempt to systematically test previous claims about the linguistic characteristics of literary naturalism and modernism. By employing quantitative statistical analyses, the present study offers a methodological alternative to previous approaches that have relied primarily on qualitative analysis. Moreover, its emphasis on discourse-level features highlights the critical importance of examining language beyond the sentence level, which is a perspective that is essential for capturing the stylistic distinctions that characterize literary texts.

Despite its contributions to the fields of corpus linguistics and corpus stylistics, this study has several limitations.

A primary limitation lies in its methodological approach, which treats *Dubliners* and *Portrait* as single, unified corpora, without considering potential stylistic variations across individual chapters or stories. While this approach allows a macro-level comparison of overall stylistic tendencies, it may obscure intra-textual fluctuations or shifts in narrative style. Future research would benefit from a more detailed analysis at the chapter or story level to capture such internal variation.

In addition, the study's focus on only two texts by a single author, albeit representative of different literary movements, limits the generalizability of the findings. In order to strengthen and broaden the empirical basis for distinguishing the stylistic features of naturalism and modernism, subsequent studies should include a wider range of literary works authored by different writers within each movement.

Another limitation of this study is the potential influence of narrative perspective and protagonist age on the linguistic patterns examined. The works included in the analysis do not employ a uniform narrative perspective or focalization. For instance, several stories in *Dubliners* adopt narrative perspectives and forms of focalization associated with child protagonists, whereas *Portrait* employs a substantially different narrative perspective and centers on a protagonist whose age and development change over the course of the novel. Such differences may directly affect a range of linguistic indices examined in this study. Consequently, some of the linguistic patterns identified as distinguishing Naturalist and Modernist texts may also be attributable, at least in part, to differences in narrative perspective, focalization, or protagonist age. These alternative explanations should therefore be taken into account when interpreting the findings. In particular, the linguistic differences observed in the present study should not be understood as straightforward evidence of stylistic distinctions between Naturalism and Modernism. Rather, the findings should be interpreted cautiously, with the possibility that features of narrative structure and characterization may contribute to the observed linguistic variation. Future research could address this limitation by controlling for or systematically examining the effects of narrative perspective, focalization, and protagonist age across a more carefully matched set of texts.

References

- Abrams, M. H. 1993. *A Glossary of Literary Terms*. Fort Worth, TX: Harcourt Brace College Publishers.
- Bosseaux, C. 2004. Point of view in translation: A corpus-based study of French translations of Virginia Woolf's *To the Lighthouse*. *Across Languages and Cultures* 5(1), 107-122.
- Can, F. and J. M. Patton. 2004. Change of writing style with time. *Computers and the Humanities* 38(1), 61-82.
- Clee, D. G. 1965. *The Theme of Class in James Joyce's Dubliners*. Doctoral dissertation, University of British Columbia, Vancouver, Canada.
- Coleman, B. G. 1991. *Imagination, Illusion and Vision in James Joyce's Dubliners*. MA thesis, McMaster University, Hamilton, Canada.
- Coltheart, M. 1981. The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology: Section A* 33(4), 497-505.
- Duran, N. D., P. M. McCarthy, A. C. Graesser and D. S. McNamara. 2007. Using temporal cohesion to predict temporal coherence in narrative and expository texts. *Behavior Research Methods* 39(2), 212-223.
- Fischer-Starcke, B. 2009. Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis. *International Journal of Corpus Linguistics* 14(4), 492-523.
- Graesser, A. C., N. Dowell and C. Moldovan. 2011. A computer's understanding of literature. *Scientific Study of Literature* 1(1), 24-33.

- Graesser, A. C. and D. S. McNamara. 2011. Computational analyses of multilevel discourse comprehension. *Topics in Cognitive Science* 3(2), 371-398.
- Graesser, A. C., D. S. McNamara, M. M. Louwerse and Z. Cai. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers* 36(2), 193-202.
- Gibbons, L. 2000. 'Have you no homes to go to?': Joyce and the politics of paralysis. In D. Attridge and M. Howes, eds., *Semicolonial Joyce*, 150-171. Cambridge: Cambridge University Press.
- Halliday, M. A. K. and R. Hasan. 1976. *Cohesion in English*. London: Longman.
- Hardie, A. and I. van Dorst. 2020. A survey of grammatical variability in Early Modern English drama. *Language and Literature* 29(3), 275-301.
- Herrmann, J. B., C. Schöch and K. van Dalen-Oskam. 2015. Revisiting style, a key concept in literary studies. *Journal of Literary Theory* 9(1), 25-52.
- Holman, C. H. and W. Harmon. 1986. *A Handbook to Literature*. New York: Macmillan.
- Ikeo, R. 2019. 'Colloquialization' in fiction: A corpus-driven analysis of present-tense fiction. *Language and Literature* 28(3), 280-304.
- Johnson, J. H. 2016. A comparable comparison? A corpus stylistic analysis of the Italian translation of Julian Barnes' *Il Senso di una Fine* and the original text *The Sense of an Ending*. *Language and Literature* 25(1), 38-53.
- Joyce, J. 2008a. *A Portrait of the Artist as a Young Man*. Oxford: Oxford University Press.
- Joyce, J. 2008b. *Dubliners*. Oxford: Oxford University Press.
- Jurafsky, D. and J. H. Martin. 2008. *Speech and Language Processing: An Introduction to Speech Recognition, Computational Linguistics and Natural Language Processing*. Upper Saddle River, NJ: Prentice Hall.
- Just, M. A. and P. A. Carpenter. 1987. *The Psychology of Reading and Language Comprehension*. Boston, MA: Allyn and Bacon.
- Kim, S. H. 2003. Eveline: A naturalist aspect. *James Joyce Journal* 9(2), 25-42.
- Kim, Y. J. and R. A. Cribbie. 2018. ANOVA and the variance homogeneity assumption: Exploring a better gatekeeper. *British Journal of Mathematical and Statistical Psychology* 71(1), 1-12.
- Kintsch, W. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95(2), 163-182.
- Lee, H. and M. Jeon. 2019. A corpus stylistic analysis of Jonathan Swift's writing style in *Gulliver's Travels* and *A Tale of a Tub*. *Korean Journal of English Language and Linguistics* 19(1), 120-141.
- Lehan, R. 1994. James Joyce: The limits of modernism and the realms of the literary text. *Arizona Quarterly: A Journal of American Literature, Culture, and Theory* 50(1), 87-108.
- Leonard, G. 1994. Joyce in America: Cultural politics and the trials of Ulysses, and: Narrative con/texts in *Dubliners*. *MFS Modern Fiction Studies* 40(2), 406-411.
- Mahlberg, M. 2013. *Corpus Stylistics and Dickens's Fiction*. London: Routledge.
- Mukhitdinova, N. 2019. Modernist features in *A Portrait of the Artist as a Young Man* by James Joyce. *European Journal of Research and Reflection in Educational Sciences* 7(12), 224-227.
- Murphy, S. 2015. I will proclaim myself what I am: Corpus stylistics and the language of Shakespeare's soliloquies. *Language and Literature* 24(4), 338-354.
- Pennebaker, J. W. and L. A. King. 1999. Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology* 77(6), 1296-1312.
- Rao, C. R. 1992. R. A. Fisher: The founder of modern statistics. *Statistical Science* 7(1), 34-48.
- Shiach, M. (Ed.). 2007. *The Cambridge Companion to the Modernist Novel*. Cambridge: Cambridge University

Press.

- Simion, M. O. 2013. Modernist techniques in *A Portrait of the Artist as a Young Man* by James Joyce. *Annals of the Constantin Brâncuși University of Târgu Jiu, Letters and Social Sciences Series* 4, 57-61.
- Sotirova, V. 2013. *Consciousness in Modernist Fiction: A Stylistic Study*. London: Palgrave Macmillan.
- Stubbs, M. 2005. Conrad in the computer: Examples of quantitative stylistic methods. *Language and Literature* 14(1), 5-24.
- Thornborrow, J. and S. Wareing. 1998. *Patterns in Language: An Introduction to Language and Literary Style*. London: Routledge.
- Totò, G. 2014. *Italian Translations of English Stream of Consciousness: A Study of Selected Novels by James Joyce and Virginia Woolf*. Doctoral dissertation, University of Edinburgh, Edinburgh, UK.
- Wales, K. 1992. *The Language of James Joyce*. London: Macmillan.

Examples in: English

Applicable Language: English

Applicable Level: Tertiary