

The Corpus-based Multidimensional Analysis of Hemingway's *The Torrents of Spring*

Jianfeng Lian¹, Suwaree Yordchim², Cholthicha Sudmuk³, Angvarrah Lieungnapar⁴, Behrad Aghaei⁵, Muhammad Shahid Khan⁶

¹Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University, Bangkok, Thailand. E-mail: s64584952007@ssru.ac.th

²Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University, Bangkok, Thailand. E-mail:suwaree.yo@ssru.ac.th

³Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University, Bangkok, Thailand. E-mail:cholthicha.su@ssru.ac.th

⁴ Department of Humanities, Faculty of Humanities and Social Sciences, Suan Sunandha Rajabhat University, Bangkok, Thailand. E-mail:angvarrah.li@ssru.ac.th

⁵Department of Middle East Studies, University of Michigan, Ann Arbor, Michigan, USA. E-mail:aghaeib@umich.edu

⁶College of Innovation and Management, Suan Sunandha Rajabhat University, Bangkok, Thailand. E-mail: shahid.kh@ssru.ac.th

Correspondence: Suwaree Yordchim, Department of Linguistics, Graduate School, Suan Sunandha Rajabhat University, Bangkok, Thailand. E-mail:suwaree.yo@ssru.ac.th

Received: September 9, 2024

Accepted: December 31, 2024

Online Published: April 17, 2025

doi:10.5430/wjel.v15n5p232

URL: <https://doi.org/10.5430/wjel.v15n5p232>

Abstract

This study employs a corpus-based multidimensional analysis, as proposed by Biber, to investigate the linguistic features of Ernest Hemingway's first novel, *The Torrents of Spring*, compared to his other works. Through a corpus analysis of 21,709 tokens from *The Torrents of Spring* as the study corpus and 725,674 tokens from eight of his later novels as the reference corpus, the study sheds light on Hemingway's unique writing style by examining his linguistic preferences. The results show notable variations in 20 linguistic characteristics, including *The Torrents of Spring's* greater use of nouns, longer average word lengths, greater prepositional frequency, and fewer causal subordinations. These findings imply that Hemingway's approach in this early work is distinguished by emphasizing informative discourse, incorporating significant substance without prioritizing interaction.

Furthermore, the study identifies six textual dimensions based on Biber's framework, highlighting that while some dimensions show similarities across Hemingway's works, dimensions such as "Involved vs. Informational Production" and "Online Informational Elaboration" exhibit striking contrasts. The analysis demonstrates that Hemingway's first novel is marked by a significant emphasis on precise lexical choices and structured information delivery, in contrast to his later writings' more interactive and emotionally nuanced style.

By offering factual proof of Hemingway's linguistic inclinations, this study advances our knowledge of his changing writing style. It also fills a vacuum in the literature concerning the particular linguistic elements of *The Torrents of Spring*, providing new light on the intricacies of Hemingway's storytelling style. The findings highlight the value of corpus-based approaches in literary analysis and help readers understand the complex relationship between language choices and story structure in Hemingway's body of work.

Keywords: corpus-based, multidimensional analysis, Hemingway, linguistic features

1. Introduction

1.1 Introduce Hemingway's First Novel

Ernest Hemingway (hereinafter "Hemingway") was one of the most influential American authors of the 20th century and the spokesman for the "Lost Generation." Hemingway's influence on modern literature is undeniable. His spare, declarative prose style and the human condition have inspired countless writers and readers alike (Wyatt, 2017). Despite his turbulent personal life and tragic end, Hemingway remains one of the most celebrated and celebrated authors of the 20th century.

Hemingway's first novel, *The Torrents of Spring*, provoked extremely controversial reactions. Critics generally dismiss the work, deeming it far less significant than *The Sun Also Rises*, published in the same year. Even though John Dos Passos found it hilarious (Baker, 1981), he didn't want it published. Hadley Richardson, Hemingway's wife at the time, found his portrayal of Anderson "evil" (Berg, 1979). However, F. Scott Fitzgerald considered the novella a masterpiece (Mellow, 1992). Meyers (1985) stated that *The Torrents of the Spring* was inferior to his later works due to Hemingway's immature writing style. Since then, *The Torrents of Spring* has received little scholarly criticism.

Writing style refers to the author's various linguistic choices in the text. "Linguistic choices" are the selection of words to create a certain effect or tone in the text, which can be identified as the author's writing style (Isti'anah, 2013). When talking about Hemingway's writing

style in *The Torrents of Spring*, most scholars state that Hemingway prefers to use short words and simple sentences to express his ideas. For example, Zheng (2013) noted that Hemingway frequently used short words and third-person pronouns in his early writings. Loudin (2013) declared that Hemingway developed a simple, repetitive writing style in his early years (before publishing *The Torrents of Spring*). Miller (2019) discusses Hemingway's characteristic minimalist style in *The Torrents of Spring*, focusing on his simple language and concise dialogue to convey complex emotions.

However, the researchers on Hemingway's writing style in *The Torrent of the Spring* never studied the relativity of textual dimension features and specific linguistic features, which can identify a writer's writing style (Leech, 2007).

1.2 Thesis Statement of this Study

This study employs the corpus-based technique of multidimensional (hereinafter "MD") Analysis (Biber, 1988), which utilizes the novel—*The Torrents of Spring* as the basis for analysis, which identifies co-occurrence patterns, namely textual dimensions, and linguistic features based on factor analysis, to compare Hemingway's usage of linguistic features in textual dimensions with his rest novels and determine his writing style from his linguistic choices in *The Torrents of Spring*.

1.3 The Torrents of Spring

1.3.1 The Brief Introduction of the Story

The Torrents of Spring is a novel by Ernest Hemingway, published in 1926. In the novella, the character Nick Adams returns to Michigan after disillusionment in Paris and feels detached from the literary community. When he arrives, Nick reconnects with Marjorie, a strong and independent woman with whom he has a romantic history. As they spend time together, their relationship is put to the test. Marjorie expresses her feelings confidently, while Nick struggles with insecurities and fear of commitment.

The tension escalates into a heated argument in which both characters express their frustration. Marjorie challenges Nick's indecision and forces him to confront his fears of love and vulnerability. After the argument, Nick retreats into nature to process his thoughts. The tranquil Michigan landscape provides the backdrop for his introspection, allowing him to reflect on his feelings for Marjorie and his own identity. As Nick thinks deeply about his relationship, he begins to realize the importance of honesty, both with himself and with Marjorie.

1.3.2 The Related Research on the Novel

The associated research on the novel focuses mainly on character dynamics, themes of disillusionment, and nature as a symbol. For instance, on character dynamics, Johnson (2018) focused on the female character Marjorie and explores Marjorie's role and her impact on Nick's development, highlighting themes of gender and power dynamics. Themes of Disillusionment and self-discovery, Grayson (2018) examines how the theme of disillusionment shapes Nick Adams' journey through his relationship with Marjorie and his reflections on the literary world, Anderson (2020) explores Nick's internal struggles and his path toward self-awareness, illustrating the tension between personal aspirations and societal expectations. For nature as a symbol, Green (2020) analyzes how the Michigan landscape mirrors Nick's emotional state, highlighting the contrast between the calmness of nature and his internal turmoil. North (2021) Examines how the transition of seasons represents Nick's journey toward self-discovery, emphasizing the symbolic significance of spring as a time of renewal and reflection.

Hemingway's writing style has drawn heated discussion since his works were published, however, few related researches on *The Torrents of Spring*. Zheng (2013) also noted that Hemingway tended to use simple words, shorter sentences, and frequently used third-person pronouns in his early writings. pointing out that Hemingway preferred to use clear, uncomplicated language and simple words in his early writing years. However, no researchers have taken a close look at Hemingway's linguistic choices in *The Torrents of Spring*. Adila (2023) Focuses on the role of dialogue in character development and examines how Hemingway's incisive conversations reveal deeper emotional undercurrents between Nick and Marjorie. Mohan (2013) analyzes how Hemingway's economical use of language contributes to the novella's thematic depth and examines how symbols enrich the narrative while maintaining brevity

1.3.3 Multidimensional (MD) Analysis

Biber (1988) conducted a factorial analysis of 67 linguistic features and inferred seven textual dimensions to examine variation between spoken and written English. In addition, he described multidimensional analysis in detail (Biber, 1995). He analyzed the following steps: First, the frequency of each 67 linguistic features is counted and converted into a normalized frequency per 1000 words. Secondly, the factor analysis is carried out using 5-7 factors in multidimensional analysis, based on the co-occurrence of 67 linguistic features in the corresponding dimensions. Language features in each factor have a corresponding factor loading that has positive and negative weights (the absolute value is between 0 and 1). The features with large absolute values (e.g., those greater than 0.35) should be retained, while the language features with relatively low values that do not contribute sufficiently to the factor are discarded.

Biber (1988) omitted the seventh dimension and nine linguistic features in total, due to the weak interpretation. Biber (1988) conducted a factorial analysis of 58 linguistic decisions and arrived at 6 textual dimensions.

(1) Textual dimension features

1) Dimension 1, Involved versus Informational Production. This dimension illustrates discourse with interactional, affective, and involved purposes versus discourse with highly informational purposes, carefully crafted and highly edited. Low dimension scores mean informational density, whereas high scores imply effective, interactional.

2) Dimension 2, Narrative versus Non-Narrative Concerns. This dimension differentiates discourse with primary narrative purposes from discourse with non-narrative purposes. The higher the scores, the higher the writer's narrative concern.

3) Dimension 3, Explicit versus Situation-Dependent Reference. Distinguish between a discourse that fully and explicitly identifies the referent through relativization and a discourse that relies on non-specific deictics and reference to an external situation for identification purposes. Lower values indicate dependence on context, whereas high values indicate independence from context.

4) Dimension 4, Overt Expression of Persuasion. These features focus on the expression of the speaker's point of view or on argumentative styles intended to persuade the addressee. The higher the score, the more clearly the author's point of view, certainty, and probability are highlighted.

5) Dimension 5, Abstract versus Non-Abstract Information. Distinguish between texts with a strong abstract and technical information focus and texts with non-abstract focuses. The higher the score, the higher the level of technical and abstract information.

6) Dimension 6, Online Informational Elaboration. These features distinguish between information discourses that are produced under highly constrained conditions and in which the information is presented in a relatively loose and fragmented manner, and other types of discourses, whether it is an information discourse that is highly integrated or a discourse, which is not informative. High values mean that the information expressed was created under specific time constraints.

Textual Dimension features are decided by dimension scores, which is the sum of the z-scores of all the linguistic features in the related dimensions. The formula of the z-score for each linguistic feature is:

$$Z_x = \frac{x - \mu_B}{\sigma_B}$$

X is the relative frequency of a feature in the user's input.

Z_x is the resulting z-score of the feature in consideration.

μ_B is the mean frequency for that feature in Biber's (1988) corpus, and σ_B is the standard deviation of that feature in Biber's (1988) corpus.

Therefore, according to the formula, Z_x depends on the difference between the relative frequency of a user's input and the mean frequency for that feature in Biber's (1988) corpus, which can be a positive value or a negative value.

(2) Specific linguistic features

9 linguistic features were deleted by Biber (1988) due to weak factor loading. These are predicative adjectives, gerunds, concessional subordination, downtoners, present-participial WHIZ deletions, existential there, appearance/appearance, the relative clause to subject positions, and split infinitives. Therefore, six dimensions and 58 linguistic features were adopted in the MD analysis (1988). As shown in Table 1.

Table 1. 58 linguistic features in Biber's framework (1988)

dimension	description	features
1. Involved vs. informational production	Low scores mean informational density, whereas high scores imply effective, interactional.	Involved production features: private verbs, that-deletions, contractions, present tenses, 2nd person pronouns, do as pro-verb, analytic negations, demonstrative pronouns, emphatics, 1st person pronouns, pronoun it, be as the main verb, causative subordinations, discourse particles, indefinite pronouns, hedges, amplifiers, sentence relatives, wh questions, possibility modals, non-phrasal coordination, wh clauses, stranded prepositions. (23) Informational production features: nouns, average word length, prepositions, type/token ratio, attributives adjectives. (5)
2. Narrative vs. non-narrative concerns	The higher the scores, the higher the narrative concern	Narrative concerns feature pat tenses, 3rd person pronouns, perfect aspect, public verbs, synthetic negations, and present participial clauses. (6)
3. Explicit versus situation-dependent reference	Lower scores indicate dependence on the context, whereas high scores mean independence from the context.	Explicit reference features: wh relative clauses on object position, pied-piping relatives, wh relative clauses on subject position, phrasal coordination, nominalizations. (5) Situation-dependent reference features: time adverbials, place adverbials, and general adverbs. (3)
4. Overt expression of persuasion	The higher the scores, the more explicitly marks the writer's point of view, certainty, and likelihood.	Overt expression of persuasion features: infinitives, predicative modals, suasive verbs, conditional subordinations, necessity modals, split auxiliaries. (6)

5. Abstract versus non-abstract information	The higher the scores, the higher the degree of technical and abstract information.	Abstract information features: conjuncts, agentless passives, past participial clauses, by passives, past participial WHIZ deletion relatives, and other adverbial subordinators. (6)
6. Online informational elaboration	High scores imply that the information expressed is produced under certain time constraints.	Online informational elaboration features: That clause as verb complements, demonstratives, <i>that</i> relative clauses on object position, that clauses as adjective complements. (4)

From the 6 dimensions and 58 linguistic features, Biber's (1988) MD framework is carried out to identify the writing styles and to improve students' writing skills. The findings do offer insights for teachers in their teaching process. However, little related research has been done on literary works. In this paper, Biber's (1988) MD framework is used to identify Hemingway's writing style by his linguistic choices.

1.4 Corpus-Based Techniques

Corpus-based techniques refer to Various computational and statistical methods used to analyze the linguistic and textual features present in a corpus, such as frequency analysis, collocation analysis, concordance analysis, stylometric analysis, and sentiment analysis (Sinclair,1991) The application of corpus-based techniques includes linguistic research, literary analysis, language learning, natural language processing, and content analysis.

Some common corpus-based techniques include:

- 1) Concordance: Concordance involves examining the occurrence of certain words or phrases within a corpus and analyzing their context. It helps identify patterns of word usage, collocations, and semantic associations.
- 2) Word frequency analysis: The technique involves counting the frequency of words or terms in a corpus to identify the most common or important lexical items. It can provide insights into the use and distribution of vocabulary.
- 3) Collocation analysis: the technique focuses on identifying and analyzing the co-occurrence patterns of words within a corpus. It helps identify words that commonly occur together and can indicate collocation preferences in a language.
- 4) Keyword analysis: Keyword analysis involves identifying words or terms that are statistically significant within a corpus compared to a reference corpus. It helps identify a specific corpus's key themes, topics, or distinctive features.
- 5) Dispersion: dispersion shows where the examined words occur in the corpus, related to individual files. dispersion can occur within a corpus file (beginning, middle, end) or across the files of a specific corpus.

Corpus-based techniques provide researchers with quantitative and data-driven insights into language usage and patterns. Corpus-based techniques are widely used in critical discourse analysis to reduce bias, uncover hidden meanings, and develop critical language awareness (Lynne,2012), which are most important in linking the recurring patterns in text with social context.

Corpus-based techniques provide researchers with quantitative and data-driven insights into language use and patterns. Corpus-based techniques are widely used in identifying linguistic choices to reduce bias (Lynne, 2012), which played the most important role in linking the recurring patterns in the text the social context.

1.5 Research Objectives

1. To compare Hemingway's usage of linguistic choices in textual dimensions in *The Torrents of Spring* with the rest of his novels.
2. To determine Hemingway's writing style from his linguistic choices in *The Torrents of Spring*.

2. Method

2.1 Research Method

This study employed a corpus-based approach by utilizing the MD analysis proposed by Biber (1988).

2.2 Research Process

- 1) Collect data from Hemingway's first novel—*The Torrents of Spring* (hereinafter "H1") as the study corpus, and the rest of his 8 novels (Hereinafter "HR") as the reference corpus. Downloading the PDF files from Library (Library Genesis. (n.d.)).
- 2) The Antfile converter software (2.0.2) is used to transform novels PDF into "txt" files. All the "txt" files are cleaned manually, to cut the acknowledgments and advertisements to get clean data, only the titles and contents are left in the "txt" files.
- 3) The software MD Analyzer Tagger (MAT 1.3.3) is used to tag all the linguistic features and compute the z-scores of 6 dimensions and 58 linguistic features in the texts.
- 4) The statistical package software is used to conduct variance analysis of the above z-scores to find the significantly different linguistic choices (textual dimensions and linguistic features) in a study corpus and a reference corpus.
- 5) Determine Hemingway's writing style in H1 from his linguistic choices in textual dimensions.

2.3 Data Collection

There are two groups of data in the study:

- 1) H1 is collected as a study corpus. The study corpus comprises 21,709 tokens and is divided into 8 parts for convenient comparison.
- 2) HR is all collected as a reference corpus. The total tokens in HR (725,674 tokens), as shown in Table 2.

Table 2. HR as a reference corpus

Name of Novels	Tokens	Publishing year
<i>The Sun Also Rises</i>	68,970	(1926)
<i>A Farewell to Arms</i>	108,756	(1929)
<i>To Have and Have Not</i>	58,808	(1937)
<i>For Whom the Bell Tolls</i>	147,465	(1940)
<i>Across the River and into the Trees</i>	69,696	(1950)
<i>The Old Man and the Sea</i>	26,736	(1952)
<i>Islands in the Stream</i>	176,960	(1970)
<i>The Garden of Eden</i>	68,283	(1986)

2.4 Data Analysis

This study follows Biber’s framework (1988), in which 6 dimensions and 58 linguistic features are used as research variables to conduct a Comparative Analysis using the software Multidimensional Analyzer Tagger (MAT 1.3.3).

1) Textual Dimensions

The 6 textual dimensions consist of (1) Informational versus Involved Production, (2) Narrative versus Non-Narrative Concerns, (3) Explicit versus Situation-Dependent Reference, (4) Overt Expression of Persuasion, (5) Abstract versus Non-Abstract Information, and (6) Online Informational Elaboration.

2) Linguistics Features

The 58 linguistic features consist of 28 linguistic features in dimension 1, 6 linguistic features in dimension 2, 8 linguistic features in dimension 3, 6 linguistic features in dimension 4, 6 linguistic features in dimension 5, and 4 linguistic features in dimension 6.

3. Result

By utilizing the Corpus-Based Approach with the MD analysis to compare Hemingway’s usage of linguistic features in textual dimensions with the rest of his novels, and determine his writing style from his linguistic choices in H1, the findings are summarized as follows:

3.1 Dimension Features between H1 and HR

The independent sampling T-test, statistics show both similarities and differences in linguistic features in textual dimensions between H1 and HR as follows:

3.1.1 The Textual Dimensions

1) The similarities

H1 and HR have some similarities in the textual dimensions, which are Dimension 2, where Hemingway focuses on narrative concern, Dimension 3, where Hemingway tends to indicate dependence on the context, Dimension 4, where Hemingway tends to show his point of view, certainty, and likelihood implicitly, and Dimension 5, Hemingway tends to use lower degree of technical and abstract information to make readers understand his novels. As it is shown in the following figure:

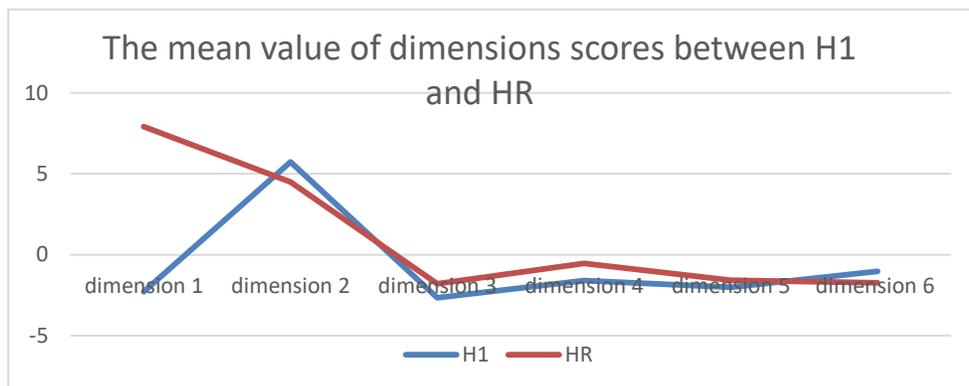


Figure 1. The mean value of dimensions scores between H1 and HR

2) The difference

There are two textual dimensions, 1 and 3, which have some significant differences. In dimension 1, H1 has a negative mean value (M=-2.28625), while HR has a positive one (M=7.9175), and the mean difference is (-10.20375), t value is (-6.016). In dimension 6, H1 and HR both have negative values. H1 (M=-1.035), and HR (M=-1.745), and the mean difference is (2.642).

The similarities and differences of linguistic features in textual dimensions between H1 and HR are shown in Table 3.

Table 3. Dimension features between H1 and HR

Dimensions	H1	HR	Mean difference	T value	P value
	Mean value	Mean value			
Dimension 1	-2.28625	7.9175	-10.20375	-6.016	0.000
Dimension 2	5.73625	4.48875	1.2475	1.441	0.172
Dimension 3	-2.6675	-1.79	-0.8775	-1.473	0.163
Dimension 4	-1.60625	-0.545	-1.06125	-1.235	0.237
Dimension 5	-2.0125	-1.575	-0.4375	-1.522	0.15
Dimension 6	-1.035	-1.745	0.71	2.642	0.019

According to the table, there are 4 dimensions (they are dimensions 2,3,4,5), which are not significantly different between H1 and HR. 2 dimensions are significantly different, they are dimensions 1 and 6.

In dimension 1, According to Biber(1988), Low scores mark high informational density and exact informational content, whereas high scores imply interactive, affective, and involved writing. In dimension 1, the mean score of H1 is -2.28625, while the dimension score of HR is 7.9175, the mean difference amounts to -10.20375. Therefore, in H1, Hemingway tends to carefully edit his first novel, enabling himself to be precise in lexical choice and an integrated textual structure. In the rest of his novels, Hemingway tends to use affective, interactional, and generalized content to build a closer tie to the readers and concern about the readers.

In dimension 6, according to Biber (1995), High scores in dimension 6 imply that the information expressed is produced under certain time constraints. In Hemingway’s H1, the mean score is (-1.035), while the mean score in Hemingway's rest of novels is (-1.745). The mean scores of all of Hemingway’s novels are below zero, which means Hemingway doesn’t express his ideas or attitudes, or integrate information with a fragmented presentation in a relatively loose manner. However, compared with HR, Hemingway likes to show his stance more implicitly and integrate the information without real-time constraints in H1.

3.2 Specific Linguistic Features

1) The similarities

An Independent sampling t-test was conducted to find significantly different specific linguistic features between H1 and HR. In H1, 38 specific linguistic features are not significantly different(p>0.05), as shown in Table 4.

Table 4. 38 specific linguistic features that are not significantly different between H1 and HR

Dimensions	No significantly different Specific linguistic features between H1 and HR (p>0.05)
Dimension 1	type-token ratio, private verbs, contractions, analytic negation, demonstrative pronouns, general emphatics, pronoun <i>it</i> , be as the main verb, discourse participles, sentence relatives, direct Wh-questions, possibility modals, non-phrasal coordination, final prepositions, Wh- clause
Dimension 2	past tense verbs, third person pronouns, perfect aspect verbs, public verbs, synthetic negation, present participle clause
Dimension 3	wh relative clauses on object positions, pie piping constructions, wh relative clauses on subject position, nomination, time adverbials, place adverbs
Dimension 4	predictive modals, suasive verbs, conditional subordination, split auxiliaries
Dimension 5	conjuncts, agentless passives, by-passives, past participial clauses, past participial WHIZ deletion
Dimension 6	<i>That</i> clauses as verb complements, <i>That</i> clause as adj. complements

2. The difference

While 20 specific linguistic features in H1 are significantly different from HR.(P<0.05) As the following table illustrates.

Table 5. 20 significantly different specific linguistic features between H1 and HR

Dimension	Linguistic features	H1	HR	Mean difference	T value	P value
		mean	mean			
Dimension 1	nouns	1.8225	-0.1725	1.995	9.684	0.000
	Word length	-0.9387	-1.6562	0.7175	4.745	0.000
	preposition	-0.8025	-1.3163	0.51375	4.104	0.001
	attributive adjectives	-0.5963	-0.9925	0.39625	2.286	0.038
	subordinator that deletion	-0.19	0.335	-0.525	-2.738	0.016
	present tense verb	-1.53	-1.001	-0.52875	-3.79	0.002
	2nd person pronouns	0.0938	0.8213	0.7275	-3.157	0.007
	do as pro-verb	-0.4863	-0.25	-0.23625	-2.563	0.023
	1st person pronouns	-0.2587	0.5113	-0.77	-3.707	0.002
	causative subordination	-0.5237	-0.1187	-0.405	-5.002	0.000
	indefinite pronouns	-0.3875	-0.0937	-0.29375	-3.472	0.004
	general hedges	-0.24	0.0863	-0.32625	-2.753	0.016
	amplifiers	-0.65	-0.0175	-0.6325	-3.332	0.005
Dimension 3	phrasal coordination	0.2138	0.8938	-0.68	-2.693	0.018
	adverbs	-1.58	-1.0137	-0.5662	-2.728	0.016
Dimension 4	infinitives	-0.8812	-0.3125	-0.56875	-2.32	0.036
	necessity modals	-0.56	0.1663	-0.7625	-3.353	0.005
Dimension 5	other adverbial subordinators	-0.1475	0.4875	-0.635	-2.269	0.04
	demonstratives	-0.2675	-1.0238	0.75625	3.555	0.003
Dimension 6	<i>That</i> clauses on object positions	-0.4775	-0.0563	-0.42125	-3.056	0.009

As the table shows 20 specific linguistic features in H1 are significantly different from HR, 13 in dimension 1, 2 in dimension 3, 2 in dimension 4, 1 in dimension 5, and 2 in dimension 6.

In dimension 1, Hemingway tends to use more nouns, longer word length, more prepositions, and more attributive adjectives, while using fewer subordinators that deletion, present verbs, 2nd person pronouns, do as pro-verb, 1st person pronouns, causative subordination, indefinite pronouns, general hedges, amplifiers.

In dimension 3, Hemingway tends to use fewer phrasal coordination and adverbs.

In dimension 4, Hemingway tends to use fewer infinitives and necessity modals.

In dimension 5, Hemingway tends to use fewer other adverbial subordinators.

In dimension 6, Hemenway tends to use more demonstrative and fewer *That* clauses on object positions.

3.3 Hemingway's Writing Style in H1 Is Determined by Linguistic Choices

3.3.1 Hemingway's Linguistic Choices in H1 in the Textual Dimension

Two of Hemingway's textual dimensions in H1, 1, and 6 exhibit different dimension features in use compared with HR, as the results of the MD analysis in Table 4 demonstrate. Dimension 1, Involved vs. Information Production, demonstrates that Hemingway represents his works with information production in H1, attempting to provide readers with the information and being more concerned with the information, by using more nouns, attributive adjectives, and longer word lengths. However, in dimension 6, online information elaboration, he used fewer demonstratives and *that* clause on object positions. It demonstrates that, in contrast to HR, Hemmingway prefers to convey his opinions subtly and to produce material without regard to time constraints.

3.3.2 Hemingway's Writing Style in H1 Is Determined by His Linguistic Choices

More nouns, longer word lengths, more prepositions, and fewer causative subordinations are the four most significant differences in Hemingway's linguistic feature usage, according to the MD analysis results. These differences can be used to identify his writing style in the following ways:

Nouns are expected to play a role in highly informative discourse, which are the primary transporters of referential meaning in written language; so, a high noun frequency denotes a high information density. Hemingway uses a tendency to employ more nouns in H1 than HR does (MD=1.995, p=0.000). As a result, in his early works, Hemingway frequently focuses on conveying messages or information to his readers.

Longer words indicate not only a high density of information but also a very exact lexical selection that produces an accurate representation of the information content. Longer words transmit more specific, specialized meanings than shorter terms. Hemingway uses longer words on average in H1 than HR does (MD=4.745, p=0.000). Since longer words provide more precise information, Hemingway's choice of word length in HI demonstrates that his writing is informational and information-oriented rather than

purpose-driven. Hemingway attempts to provide readers with as much information as he can, as evidenced by his information-focused writing style.

Additionally useful for incorporating a lot of information into a sentence are prepositional phrases. Prepositions are grammatical constructions that express a logical connection or a spatial or temporal relationship between two or more individuals, locations, or objects. Hemingway uses a tendency to employ more prepositions in H1 than in HR (MD=4.104, $p=0.001$), which means Hemingway didn't create a colloquial writing style to aid the reader in understanding his thoughts, as evidenced by the increased use of prepositions and prepositional phrases in his first book, which makes it more instructive for readers.

Causal subordination is associated with the expression of information under real-time production constraints, when there is little opportunity for elaboration through precise lexical selection. In H1, Hemingway tends to use fewer causal subordinations compared to HR (MD=-5.002, $p=0.000$). This indicates that Hemingway took more effort in selecting words to accurately convey his ideas and information.

In sum, Hemingway frequently makes use of these four linguistic choices to incorporate a lot of information into a sentence and deliver it as clearly and concisely as he can. These tasks offer plenty of opportunity for meticulous information integration and exact lexical selection, but they also call for a high degree of information focus.

In conclusion, Hemingway's language selections in the first novel indicate more information output than intentionality. His informative writing style is distinguished by a high noun density, longer words, more prepositions, and fewer causal subordinations.

4. Discussion

Compared to his later novels, Hemingway's early work displays information in a less fragmented manner, according to the results of the "Online Informational Elaboration" dimension. This implies a conscious decision to preserve accuracy and clarity in his early story, which can be indicative of his stage of development as a writer. The goal to deliver specific and in-depth information is suggested by the use of larger words and a wider range of nouns.

Beyond Hemingway's writing style, the research's findings shed light on the larger environment of early 20th-century American literature. The intricate language elements found in *The Torrents of Spring* might reflect the literary and cultural climate of the day when writers were delving deeper and deeper into the complexity of the human condition. By placing Hemingway within this framework, we can better comprehend how his early works prepared the way for the stylistic advances evident in his later masterpieces.

This study also shows how useful corpus-based approaches are for literary analysis and how they can produce quantitative evidence that supports qualitative interpretations. This method not only improves our comprehension of Hemingway's writing but also encourages future scholars to use comparable approaches with other authors and genres. By using these approaches, researchers might identify patterns and trends that could have gone overlooked in the past, leading to a more thorough knowledge of the development of literature.

In summary, the study's findings greatly advance the body of knowledge on Hemingway by highlighting the distinctive linguistic traits that characterize his early writing. Using a thorough corpus-based analysis, we have shown how Hemingway's language decisions in *The Torrents of Spring* reflect his larger narrative goals and stylistic development.

5. Conclusion

This study has explored the linguistic features of Ernest Hemingway's *The Torrents of Spring* through a corpus-based multidimensional analysis, revealing significant insights into his writing style in this early novel compared to his later works. Utilizing Biber's framework, we identified key differences in linguistic choices across six textual dimensions, providing a nuanced understanding of how Hemingway's prose evolved.

Through a corpus-based multidimensional analysis, this study has examined the linguistic choices of Ernest Hemingway's *The Torrents of Spring*, providing important new information on his writing style in this early book as opposed to his subsequent works. By applying Biber's framework, we could discern significant variations in word choices along six textual dimensions, which gave us a more sophisticated comprehension of how Hemingway's prose changed over time.

The research emphasized that *The Torrents of Spring* has a more informative and content-driven approach, as evidenced by the greater usage of prepositions, longer word lengths, and a higher frequency of nouns. This starkly contrasts the features of his later novels, which tend to include more engaging and emotionally charged content. The results imply that Hemingway's early writing was more concerned with providing information in an organized way than it was with interacting with the reader.

The analysis's most notable finding is the distinction between the "Involved vs. Informational Production" dimension. His later works have a positive mean score, suggesting a change towards an interactive style aimed at creating a connection with readers, while *The Torrents of Spring* exhibits a negative mean score, indicating a preference for informational density.

Acknowledgments

We greatly appreciate the valuable contributions of our community advisory committee members at Suan Sunandha Rajabhat University. We would also like to thank our team members who took the time to participate in this study.

Authors' contributions

Dr. Cholthicha Sudmuk and Dr. Suwaree Yordchim were responsible for the study design and revising. Dr. Angvarrah Lieungnapar was responsible for data collection. Jianfeng Lian drafted the manuscript and Prof. Behrad Aghaei and Dr. Muhammad Shahid Khan revised it. All authors read and approved the final manuscript.

Funding

Graduate School, Suan Sunandha Rajabhat University supported this work.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

Open access

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

References

- Adila, F. F. P. (2023). *The main characters' safety and love belonging needs in Ernest Hemingway's novel The Torrents of Spring* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- Anderson, M. (2020). The quest for identity: self-discovery in *The Torrents of Spring*. *Journal of American Literature*.
- Baker, C. H. (1981). *Ernest Hemingway: Selected Letters 1917-1961*. New York: Charles Scribner's Sons.
- Berg, A. S. (1979). *Max Perkins: Editor of Genius*. New York: Washington Square Press.
- Biber, D. (1988). *Variation across Speech and Writing*. Cambridge University Press, Cambridge.
<https://doi.org/10.1017/CBO9780511621024>
- Biber, D. (1995). *Dimensions of Register Variation: A Cross-linguistic Comparison*. Cambridge University Press, Cambridge.
<https://doi.org/10.1017/CBO9780511519871>
- Carter, M. (2018). Dialogue and Characterization in *The Torrents of Spring*. *Studies in Fiction*, 11(1).
- Grayson, L. (2018). "Disillusionment in Hemingway's *The Torrents of Spring*". *Modernist Studies Journal*.
- Green, R. (2020). Nature as a Reflection of Emotion in *The Torrents of Spring*. *Journal of Literary Geography*.
- Isti'anah, A. (2013). *Linguistic Features for More Understanding Literary Work*. The 1st Literary Studies Conference, Yogyakarta, Indonesia. 10.
- Johnson, S. (2018). *Feminine strength in Hemingway: analyzing Marjorie in The Torrents of Spring*. Isti.
- Leech, G. (2007). *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. Pearson Longman. New York.
- Library Genesis. (n.d.). *Library Genesis*. Retrieved from <https://libgen.is/>
- Loudin, Z. (2013). *Evolution of Writing Style in Ernest Hemingway's Works from 1916 to 1929*. Undergraduate thesis, Kent State University.
- Lynne, K. S. (2012). *Hemingway*. Harvard University Press.

- Mellow, J. R. (1992). *Hemingway: A Life Without Consequences*. New York: Houghton Mifflin.
- Meyers, J. (1985). *Hemingway: A Biography*. London: Macmillan. ISBN 0-333-42126-4.
- Miller A. (2019). Minimalism and Precision: The Writing Style of The Torrents of Spring. *Modernist Literary Criticism*, 12(3).
- Mohan, B. (2013). Style, art and technique in Hemingway's writings. *Asian Journal of Multidimensional Research (AJMR)*, 2(10), 7-13.
- North E. (2021). "Seasons of Self-Discovery: Nature's Role in The Torrents of Spring. *American Literary Studies*.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.
- Wyatt, D. (2017). *Hemingway, Style and the Art of Emotion*. University of Maryland. College Park.
- Zheng, M. (2013). The Shift of Narrative Perspective and Multiple Narrative Perspectives in Hemingway's Novels. *Studies in Literature and Language*, 7(1).1-6.