Text Typology: A Corpus-Based Text Linguistic Exploration of the Old and New Testaments

Samuel Oyeyemi Agbeleoba, Ph.D
Department of English and Literary Studies,b
Ekiti State University, Ado-Ekiti, Nigeria.
samuel.agbeleoba@eksu.edu.ng
https://orcid.org/0000-0002-9643-7682
Busayo Rebecca Ajayi-Owoyemi,
Institute of Technology-Enhanced learning and Digital Humanities,
Federal University of Technology, Akure, Nigeria.
brajayi@futa.edu.ng
https://orcid.org/0009-0007-7393-759X
Clement Temitope Ogunlesi, Ph.D
Department of Religions and Cultural Studies
Ekiti State University, Ado-Ekiti, Nigeria.
temitope.ogunlesi@eksu.edu.ng
https://orcid.org/0000-0003-1594-6598
Oluwabusayo Foluso Fafiyebi,
Department of English and Literary Studies,
Federal University, Oye-Ekiti, Nigeria
Oluwabusayo.fafiyebi@fuoye.edu.ng
https://orcid.org/0009-0008-6453-944X
Peter Adedayo Agbeleoba,
Department of Intercultural Studies (World Religions)
Nigerian Baptist Theological Seminary, Ogbomoso, Nigeria.
adedayoomooba@gmail.com

Abstract
This article presents a comprehensive corpus-based text linguistic exploration of the Old and New Testaments, focusing on text typology and genre classification. The biblical corpus, comprising over 750,000 words across 66 books, presents a unique challenge for linguistic analysis due to its diverse authorship, historical span, and multiplicity of text types. By applying principles of corpus linguistics, multi-dimensional analysis, and systemic functional linguistics, this study investigates the linguistic features that characterize different biblical text types, including narrative, poetry, law, prophecy, and epistolary discourse. The research integrates recent advancements in the ethnography of communication and discourse analysis to provide a robust framework for understanding how textual coherence and register variation operate within sacred texts. The findings demonstrate that a corpus-based approach to text typology not only clarifies the structural and functional dimensions of biblical genres but also enhances the interpretative accuracy of complex dialogic and narrative sequences. Furthermore, the inclusion of contemporary analytical models, such as those examining sacred speech events, illustrates the enduring relevance of linguistic methodologies in theological and biblical studies. This exploration underscores the value of empirical linguistic methodologies in bridging the gap between ancient textual forms and modern interpretative practices.
Keywords: Text Typology, Corpus Linguistics, Biblical Texts, Register Variation, Systemic Functional Linguistics, Ethnography of Communication.

1. Introduction

The Bible is arguably the most monumental and influential literary corpus in human history, profoundly shaping cultural, theological, and linguistic development across millennia. Comprising 66 books in the Protestant canon, the biblical text spans approximately 15 centuries of writing and encompasses a vast and complex array of literary genres and text types (Walton, 2019:45). The Old Testament, written primarily in Biblical Hebrew with portions in Aramaic, presents a rich tapestry of historical narratives, intricate legal codes, highly structured poetic compositions, and impassioned prophetic oracles. This ancient collection contrasts with and complements the New Testament, written in Hellenistic (Koine) Greek, which features detailed gospel narratives, historical accounts of the early church, complex epistolary correspondence, and vivid apocalyptic visions. Understanding the linguistic, structural, and functional characteristics of these diverse texts requires a rigorous methodological approach that extends far beyond the boundaries of traditional literary and historical criticism. Text typology, defined as the systematic classification of texts based on their internal linguistic features and communicative functions, offers a powerful and empirical framework for analyzing the biblical corpus. In recent decades, the advent and rapid expansion of corpus linguistics have revolutionized the study of language by enabling researchers to analyze massive collections of naturally occurring texts quantitatively and empirically (Biber, 1989:5). When applied to the field of biblical studies, corpus-based text linguistics provides essential tools for identifying hidden patterns of language use, tracking register variation across different authors and time periods, and mapping discourse structures that might otherwise remain obscured to the naked eye. This article aims to explore the text typology of the Old and New Testaments through a comprehensive corpus-based linguistic lens, examining in detail how different genres are constructed linguistically and how they function communicatively within their respective socio-historical contexts. The primary objective of this expansive study is to demonstrate the efficacy and necessity of corpus linguistics and multi-dimensional analysis in categorizing and interpreting biblical texts. By systematically investigating the distribution of specific linguistic features—such as pronoun usage, verb tense and aspect, lexical density, and syntactic complexity—across various traditional genres, this research seeks to elucidate the profound relationship between text form and communicative purpose. Furthermore, the article will delve deeply into how concepts such as register, cohesion, coherence, and textuality operate within specific biblical narratives, legal codes, poetic structures, and dialogic exchanges. Drawing on recent and innovative scholarship in the ethnography of communication, the study will also highlight how socio-cultural norms dictate linguistic choices in sacred speech events. Through this exhaustive exploration, the study contributes significantly to the growing intersection of linguistics and biblical scholarship, offering new, empirically grounded insights into the structural and functional dynamics of sacred texts.

2. Literature Review

The intersection of modern linguistics and traditional biblical studies has garnered significant scholarly attention over the past few decades, with researchers increasingly employing empirical and computational methodologies to analyze ancient texts. This comprehensive literature review synthesizes key theoretical developments and methodological advancements in text typology, corpus linguistics, discourse analysis, and the ethnography of communication as they relate to the rigorous study of the Old and New Testaments.

2.1. Foundations of Text Typology and Multi-Dimensional Analysis

The theoretical foundation of modern, empirically driven text typology is deeply rooted in the pioneering and highly influential work of Douglas Biber. Biber's development of the multi-dimensional (MD) analysis framework revolutionized the empirical classification of texts by moving away from intuitive categorization toward statistical rigor. Biber (1989:6) developed a comprehensive typology of English texts based on the statistical co-occurrence of specific linguistic features, utilizing factor analysis to identify underlying dimensions of variation. Through this methodology, he identified distinct text types such as intimate interpersonal interaction, informational interaction, scientific exposition, learned exposition, imaginative narrative, general narrative exposition, situated reportage, and involved persuasion.

A critical component of Biber's methodology is the sharp distinction he draws between "genres" and "text types." Genres, in Biber's framework, are categorized based on external, situational criteria—such as the author's purpose, the intended audience, and the physical medium of communication (e.g., a novel, a newspaper article, a personal letter). In contrast, "text types" are defined strictly by their internal linguistic characteristics—the actual words, grammatical structures, and syntactic patterns used within the text (Biber, 1989:39). This functional approach to text classification posits that texts belonging to a particular text type share frequent use of specific linguistic features that reflect a common, underlying communicative purpose, regardless of their situational genre.In the context of biblical studies, the application of MD analysis allows scholars to move significantly beyond traditional, often intuitive genre classifications. It provides a mechanism to empirically identify the linguistic markers that define and differentiate biblical narratives, poetry, legal texts, and prophecies. The crucial distinction between situational genres and linguistic text types is particularly illuminating for understanding the complex, multi-layered nature of the biblical corpus. For example, a single biblical book, traditionally classified under a single genre heading like "Prophecy," may contain multiple distinct linguistic text types, fluidly shifting between narrative exposition, poetic lament, and involved persuasion. Recognizing this linguistic diversity within traditional genre boundaries is essential for accurate exegesis and translation.

2.2. The Rise of Corpus Linguistics in Biblical Studies

Corpus linguistics, characterized by the empirical, computer-assisted analysis of large, electronic collections of texts, has increasingly and successfully been applied to ancient languages, including Biblical Hebrew, Aramaic, and Hellenistic Greek. This methodological shift represents a move from prescriptive grammar to descriptive, usage-based linguistic analysis.Matthew Brook O'Donnell (2005:12) conducted groundbreaking and foundational research by systematically applying corpus linguistic techniques to the Greek of the New Testament. His comprehensive work demonstrated how advanced computational methods, such as collocational analysis and statistical modeling, could significantly refine lexicography, textual criticism, and discourse analysis of biblical texts. By treating the New Testament not merely as a sacred document but as a specialized linguistic corpus, O'Donnell showed that empirical data could resolve long-standing ambiguities in word meanings, highlight subtle syntactic patterns, and reveal discourse structures that traditional, manual grammatical studies often overlooked or misinterpreted. His application of these techniques to issues like synonymy and discourse annotation opened new avenues for understanding the situational context of New Testament letters.The application of corpus linguistics to the Old Testament has similarly yielded immensely valuable insights, though it presents unique challenges due to the chronological depth and complex transmission history of the Hebrew text. Researchers have increasingly utilized computational linguistics to analyze the vast variation in Biblical Hebrew, exploring complex questions related to authorship, diachronic dating, and dialectal differences. The ability to precisely quantify linguistic features across the extensive Old Testament corpus—which comprises over 580,000 words in standard English translations—enables a much more objective and rigorous assessment of stylistic shifts. For instance, corpus studies have been instrumental in mapping the linguistic transition from Standard Biblical Hebrew (SBH), typical of pre-exilic texts, to Late Biblical Hebrew (LBH), characteristic of post-exilic writings (Yount, 2021:88). These empirical findings provide crucial data points for historical-critical debates regarding the dating and composition of various Old Testament books.

2.3. Discourse Analysis, Systemic Functional Linguistics, and Textuality

Discourse analysis and the study of textuality focus on the macro-level structures of language—how individual sentences and utterances cohere to form larger, meaningful texts, and how these texts function within specific social contexts. Systemic Functional Linguistics (SFL), developed by M.A.K. Halliday and refined by numerous scholars, has been particularly influential and productive in this area of research. SFL views language fundamentally as a social semiotic system—a resource for making meaning in social contexts—and emphasizes the inseparable role of context in shaping linguistic choices.A key component of the SFL framework is register analysis, which examines texts in terms of three contextual variables: field, tenor, and mode. 'Field' refers to the subject matter and the nature of the social action taking place; 'tenor' describes the relationship between the participants, including their roles, status, and power dynamics; and 'mode' refers to the channel of communication (e.g., written vs. spoken) and the rhetorical mode (e.g., persuasive, didactic) (Halliday & Hasan, 1989:12).

In contemporary biblical scholarship, register and discourse analysis have been fruitfully employed to interpret complex prophetic, poetic, and narrative texts. For instance, Colin Toffelmire (2016:28) utilized the SFL framework to conduct a meticulous discourse and register analysis of the prophetic book of Joel. His study demonstrated compellingly how the specific linguistic choices made by the author reflect and construct the socio-semiotic context of ancient Israel, shedding new light on the book's structure and theological message. Similarly, extensive studies on cohesion and coherence in the Old Testament have highlighted how biblical authors utilized sophisticated linguistic devices—such as lexical repetition, strategic use of conjunctions, chiasmus, and structural parallelism—to create highly unified, memorable, and persuasive texts (Berlin, 1989:22).

2.4. Ethnography of Communication in Sacred Texts

More recently, biblical scholarship has begun to integrate methodologies from linguistic anthropology, specifically the ethnography of communication, into biblical discourse analysis. This approach views speaking and writing not merely as the transmission of information, but as culturally situated social actions governed by specific norms and rules.

A highly notable and recent example of this integrative approach is the comprehensive study by Agbeleoba et al. (2026:58), which undertook an in-depth ethnographic and textual analysis of the conversation between Jesus Christ and the thieves on the cross, as portrayed in the Synoptic Gospels. Utilizing Dell Hymes's highly influential SPEAKING model—which analyzes Setting, Participants, Ends, Act sequences, Key, Instrumentalities, Norms, and Genre—alongside rigorous principles of textuality, Agbeleoba et al. (2026:60) examined how communicative actions within this specific, highly charged sacred speech event were shaped by the socio-cultural contexts of first-century Judea and structured through linguistic coherence.

Their meticulous analysis revealed how the brief dialogue functions as a dense, unified text that reflects deep socio-religious norms regarding judgment, honor, and salvation, highlighting the sharply contrasting intentions and theological postures of the repentant and impenitent thieves. This integrative approach powerfully demonstrates the effectiveness of combining ethnographic frameworks with precise textual analysis to unpack the complex pragmatic, social, and theological dimensions of biblical dialogues (Agbeleoba et al., 2026:64). It underscores that biblical texts are not just repositories of doctrine, but records of dynamic social interactions that require culturally sensitive linguistic analysis to be fully understood.

3. Methodology

To achieve a comprehensive and robust understanding of the text typology of the Old and New Testaments, this study adopts a rigorous, multi-tiered corpus-based text linguistic methodology. The methodological framework carefully integrates quantitative principles from corpus linguistics and multi-dimensional analysis with qualitative insights from systemic functional linguistics and the ethnography of communication. This hybrid approach ensures that the analysis is both empirically grounded and contextually nuanced.

3.1. Corpus Compilation, Preparation, and Annotation

The foundational step in this research involves the compilation and preparation of the biblical corpus. While acknowledging the primacy of the original Hebrew, Aramaic, and Greek texts, this study utilizes a highly regarded English translation as the primary data source to facilitate broad typological comparison and to make the findings accessible to a wider interdisciplinary audience. The English Standard Version (ESV) was selected for its recognized balance of literal, word-for-word accuracy (formal equivalence) and contemporary syntactic readability, making it highly suitable for structural linguistic analysis.

The total biblical corpus analyzed comprises exactly 756,846 words. This massive dataset is distributed across the 39 books of the Old Testament, which contain 581,112 words, and the 27 books of the New Testament, which contain 175,734 words. To facilitate systematic comparative analysis, the entire corpus was meticulously segmented into major genre categories, reflecting traditional classifications widely recognized in biblical scholarship:

Law/Pentateuch (e.g., Genesis, Exodus, Leviticus, Numbers, Deuteronomy) - Characterized by foundational narratives and extensive legal/ritual codes.

Historical Narrative (e.g., Joshua, Judges, 1 & 2 Kings, 1 & 2 Chronicles, Acts) - Focused on chronological accounts of national and early church history.

Wisdom and Poetry (e.g., Psalms, Proverbs, Job, Ecclesiastes, Song of Solomon) - Marked by poetic structure, parallelism, and philosophical reflection.

Prophecy (e.g., Isaiah, Jeremiah, Ezekiel, Daniel, Minor Prophets) - Containing divine oracles, visions, and hortatory discourse.

Gospel Accounts (e.g., Matthew, Mark, Luke, John) - Theological biographies featuring high concentrations of narrative and reported speech.

Epistolary Discourse (e.g., Romans, 1 & 2 Corinthians, Galatians, Hebrews) - Didactic and persuasive letters addressed to specific communities or individuals.

Apocalyptic Literature (e.g., Revelation, portions of Daniel) - Highly symbolic, visionary texts dealing with eschatological themes.

The texts were digitised, cleaned, and heavily annotated to facilitate advanced computational analysis. Comprehensive metadata tags were applied to every segment of the text to indicate the specific book, chapter, verse, traditional genre classification, and, crucially for dialogic sections, the identified speaker and addressee. This highly structured, annotated corpus allows for the rapid and precise extraction of quantitative data regarding word frequencies, syntactic patterns, part-of-speech distributions, and complex lexical collocations across different sections of the Bible.

3.2. Analytical Framework: A Three-Phased Approach

The linguistic analysis proceeds systematically through three interconnected and mutually reinforcing phases:

Phase 1: Quantitative Feature Extraction and Profiling   Utilizing specialized corpus analysis software (such as AntConc or WordSmith Tools), the study first extracts vast amounts of quantitative data regarding the distribution of specific linguistic features across the various biblical sub-corpora. Drawing heavily on the methodological foundations laid by Biber (1989:14), the analysis focuses on identifying and quantifying features that are known to be strong indicators of register variation and text type. These features include, but are not limited to:

- Pronoun distribution: The ratio of first-person (I, we) and second-person (you) pronouns (indicating interpersonal involvement) versus third-person (he, she, it, they) pronouns (indicating narrative distancing).
- Verb tense and aspect: The frequency of past tense (typical of narrative) versus present and future tenses (typical of exposition and prophecy).
- Syntactic complexity: The use of coordinating conjunctions (and, but) versus subordinating conjunctions (because, although, if), which indicates the level of structural elaboration.
- Lexical density and diversity: The ratio of content words (nouns, verbs, adjectives) to grammatical words (prepositions, articles), providing a measure of informational density.
- Passive voice constructions: Indicating a focus on the action or the receiver of the action rather than the agent, often used in legal or objective reporting.

Phase 2: Multi-Dimensional Text Typology and Clustering   In the second phase, the extracted linguistic features are subjected to rigorous statistical analysis, specifically multivariate techniques such as factor analysis and hierarchical cluster analysis, to identify significant co-occurrence patterns. The goal is to observe which linguistic features consistently group together across different texts. Texts that cluster tightly together based on their internal linguistic characteristics are categorized into distinct "linguistic text types."

Crucially, these empirically derived text types are then mapped against their traditional, situational genre classifications. This comparative phase aims to reveal the underlying functional dimensions of biblical language. For example, it seeks to determine whether a text traditionally labeled "Prophecy" clusters linguistically with "Poetry" (due to high structural parallelism and figurative language) or with "Epistolary Discourse" (due to high interpersonal involvement and hortatory features). This phase exposes the true linguistic diversity hidden beneath broad genre labels.

Phase 3: Qualitative Discourse, Textuality, and Ethnographic Analysis   To complement, contextualize, and interpret the quantitative findings from the first two phases, a rigorous qualitative discourse analysis is applied to selected, representative passages from each identified text type. Utilizing the Systemic Functional Linguistics framework, this phase examines the 'field', 'tenor', and 'mode' of specific texts to understand how they function communicatively within their contexts. Furthermore, the analysis evaluates the principles of textuality—specifically cohesion (how the text is glued together syntactically) and coherence (how the text makes logical and semantic sense).

For dialogic and interactive texts, this phase incorporates ethnographic models, specifically Hymes's SPEAKING framework, to systematically unpack the complex socio-cultural dynamics, power relations, and unstated norms governing the speech events. This qualitative deep-dive builds directly upon the robust methodology demonstrated by Agbeleoba et al. (2026:59), ensuring that the linguistic data is always interpreted within its rich socio-historical and theological matrix.

4. Corpus Characteristics and Genre Distribution: A Macro-Level View

Before delving into specific text types, it is essential to establish a macro-level understanding of the biblical corpus. The Bible is characterized by immense structural, historical, and linguistic diversity. A quantitative overview of the corpus reveals significant and revealing variations in word count, lexical density, and genre distribution between the Old and New Testaments, reflecting their different origins and purposes.

4.1. The Old Testament Corpus: Scale and Narrative Dominance

The Old Testament is a massive collection, constituting approximately 77% of the total biblical corpus (581,112 words in the ESV). It is a sweeping, multi-generational anthology dominated by extensive historical narratives, complex legal codes, and highly structured poetic compositions. The sheer scale of the Old Testament is evident in its longest books, which dwarf most New Testament writings. For example, the Book of Jeremiah contains 40,508 words, Genesis contains 36,326 words, and the Book of Psalms contains 42,297 words.

When analysed quantitatively, the major Old Testament genres exhibit highly distinct and consistent linguistic profiles:

- Historical Narrative (e.g., Genesis, Joshua, 1 & 2 Kings): These texts form the backbone of the Old Testament. Linguistically, they are characterized by an overwhelmingly high frequency of third-person pronouns (he, they) and past tense verbs. Syntactically, they rely heavily on coordinating conjunctions, most notably the Hebrew waw consecutive (often translated simply as "and" or "then" in English), which drives the narrative forward in a relentless, sequential chain. These features prioritize chronological reporting, objective distancing, and the seamless presentation of events. Lexical density in these sections tends to be lower, as the focus is on action and sequence rather than dense description.

- Law and Legal Codes (e.g., Leviticus, Deuteronomy): Found predominantly in the Pentateuch, these texts display a starkly different linguistic profile. They are marked by a high density of imperative verbs (commands), modal verbs indicating obligation (shall, must), and complex conditional clauses ("If a man does X... then Y shall happen"). The vocabulary is highly specialized, featuring precise terminology related to ritual purity, sacrificial procedures, and covenantal obligations. The register is formal, authoritative, and prescriptive, designed to regulate community behavior and religious practice.

- Poetry and Wisdom Literature (e.g., Psalms, Proverbs, Job): These texts represent the artistic and philosophical pinnacle of the Old Testament. Linguistically, they are defined not by rhyme or meter in the English sense, but by structural parallelism—the pairing of lines that echo, contrast, or expand upon each other's meaning. They exhibit a high frequency of figurative language, metaphors, and similes. Unlike historical narratives, poetic texts show a much higher proportion of first-person (I, my) and second-person (you, your) pronouns, reflecting a deeply interpersonal, emotive, and often dialogic communicative function (e.g., the Psalmist addressing God, or Wisdom addressing the reader).

- Prophecy (e.g., Isaiah, Jeremiah, Amos): Prophetic literature is perhaps the most linguistically complex genre in the Old Testament, frequently alternating between narrative frameworks (describing the prophet's life) and direct, impassioned divine oracles. The oracles often employ elevated, highly poetic language, rich in imagery and metaphor. They are characterized by future tense constructions (predicting judgment or restoration), imperative verbs (calling for repentance), and hortatory discourse. The register shifts rapidly between the voice of the prophet and the voice of God, creating a dynamic and highly involved text type.

4.2. The New Testament Corpus: Interactivity and Persuasion

The New Testament is significantly shorter, comprising only 23% of the total corpus (175,734 words). However, it introduces entirely new genres and text types that reflect its distinct historical context—the Greco-Roman world of the first century—and its specific theological purpose: documenting the life of Jesus and guiding the early Christian communities.

- Gospel Narratives (Matthew, Mark, Luke, John): While the four Gospels share basic linguistic similarities with Old Testament historical narratives (e.g., past tense verbs, third-person perspective), they are uniquely characterized by an exceptionally high concentration of reported speech and dialogue. The Gospels are essentially dialogic narratives. Furthermore, the parables of Jesus, embedded extensively within these narratives, represent a distinct and highly effective sub-text type. Parables utilize metaphorical language, allegorical structures, and everyday, agrarian or domestic vocabulary to convey profound and often subversive theological concepts.

- Epistolary Discourse (e.g., Romans, 1 Corinthians, Galatians): The letters written by Paul, Peter, John, and others constitute a major and highly distinctive portion of the New Testament. Linguistically, these texts are intensely interactive. They are characterized by direct address, evidenced by the extremely frequent use of first-person ("I," "we") and second-person ("you") pronouns. They heavily employ rhetorical questions, imperative verbs (exhortations to specific behaviors), and complex, hypotactic argumentative structures marked by logical subordinating connectors (e.g., "therefore," "because," "so that," "in order that"). The register is highly persuasive, pastoral, and deeply embedded in the specific socio-cultural controversies of the recipient communities.

- Apocalyptic Literature (The Book of Revelation): Revelation represents a highly specialized, almost esoteric text type. While it shares some features with Old Testament prophecy, it is distinguished by its dense, overwhelming use of symbolic imagery, repetitive structural formulas (e.g., sequences of sevens), and visionary narrative techniques. The language is designed to evoke awe, terror, and ultimate hope, utilizing a register that is radically different from the logical argumentation of the epistles or the straightforward reporting of the Gospels.

4.3. Traditional Genres vs. Linguistic Text Types

When Biber's (1989:39) crucial distinction between situational genres and empirically derived linguistic text types is rigorously applied to the biblical corpus, several fascinating and complex patterns emerge. It becomes immediately apparent that traditional genre classifications, while useful for general orientation, often mask the profound internal linguistic diversity of individual biblical books.

For example, the Book of Exodus is traditionally and broadly categorized as "Law" or "Torah." However, a detailed corpus linguistic analysis reveals that it is a composite document containing multiple, highly distinct text types. The early chapters (Exodus 1-14) are classic historical narrative, detailing the oppression in Egypt and the dramatic escape. The middle chapters (Exodus 20-23) shift abruptly into legal exposition, presenting the Ten Commandments and the Book of the Covenant with their high density of imperatives and conditionals. The later chapters (Exodus 25-31) shift again into highly detailed instructional and procedural text types, outlining the precise architectural and artistic specifications for constructing the Tabernacle. Treating Exodus as a single, uniform genre ignores this massive linguistic variation.

Similarly, the major prophetic books (like Isaiah or Jeremiah) are not linguistically monolithic. They constantly blend "Imaginative Narrative" (when describing vivid visions or symbolic actions performed by the prophet) with "Involved Persuasion" (when delivering impassioned calls to repentance, ethical exhortations, or pronouncements of doom). This multi-layered textuality requires a highly nuanced analytical approach—one that recognizes how biblical authors fluidly and masterfully shifted between linguistic registers and text types to achieve their complex, multifaceted communicative goals within a single document.

5. Text Linguistic Exploration of Biblical Discourse

To fully appreciate the sophisticated text typology of the Old and New Testaments, it is necessary to move beyond macro-level statistics and examine precisely how specific linguistic features function dynamically within the broader context of biblical discourse. This section explores the detailed application of systemic functional linguistics, discourse analysis, and the ethnography of communication to key, representative biblical text types.

5.1. The Architecture of Cohesion and Coherence in Old Testament Narrative

Historical narratives in the Old Testament are not merely collections of facts; they are masterclasses in textual cohesion and narrative art. The ancient authors employed highly specific and effective linguistic devices to create a continuous, coherent, and compelling storyline across vast temporal and geographical expanses.

In Biblical Hebrew, the primary and most iconic cohesive device in narrative texts is the waw consecutive. This syntactic feature (often translated as "and," "then," or "so" in English) is prefixed to verbs and serves to link sequential actions, maintaining the relentless chronological flow of the discourse (Berlin, 1989:24). From a text linguistic perspective, this creates a text type that exhibits a relatively low degree of lexical density (fewer unique vocabulary words per total words) but an exceptionally high degree of structural coherence. The language is heavily characterized by coordination (linking clauses of equal weight) rather than subordination (embedding dependent clauses), which creates a steady, rhythmic, and highly readable pacing suitable for oral recitation.

Beyond syntax, Old Testament narratives rely heavily on lexical cohesion to build thematic depth. For example, in the extended narrative of the Exodus, the text utilizes repetitive lexical chains and motif words. Phrases like "let my people go," "serve me," and the repeated noting that Pharaoh "hardened his heart" (or that God hardened it) act as structural pillars. These repetitions do not merely link sentences; they build profound thematic coherence, continually underscoring the central theological tension between divine sovereignty and human stubbornness.

Register analysis of these narratives reveals a highly specific and consistent tenor: the narrator almost universally adopts an omniscient, objective, and deeply authoritative stance. The narrator rarely intrudes into the text with first-person commentary or explicit moralizing. This deliberate linguistic distancing serves a crucial rhetorical function: it elevates the authority of the text, presenting the recorded events not as personal reflections or subjective opinions, but as absolute, factual, historical realities dictated by divine providence.

5.2. The Pragmatics and Persuasion of Epistolary Discourse

In stark contrast to the objective, distanced tenor of Old Testament historical narratives, the New Testament epistles are characterized by intense, urgent interpersonal involvement. A corpus-based analysis of the Pauline epistles, for instance, reveals a linguistic profile that aligns almost perfectly with what Biber (1989:22) identifies as the "Involved Persuasion" text type.

The most prominent feature is the exceptionally high frequency of first-person ("I," "we") and second-person ("you," "brothers") pronouns. The author is constantly foregrounding his relationship with the audience. Furthermore, the epistles are saturated with rhetorical questions, which serve to engage the reader's mind and anticipate objections, and imperative verbs, which directly command specific ethical behaviors or theological shifts.

The textuality of the epistles is highly complex, often seamlessly blending dense theological exposition with urgent practical exhortation. Cohesion in these texts is achieved not primarily through chronological sequence (like the waw consecutive), but through logical and argumentative connectors. Words like "therefore" (oun), "for" (gar), "because" (hoti), and "so that" (hina) are the glue that holds the epistles together, structuring the complex, often convoluted argumentative sequences that characterize Paul's writing.

The register of the epistles is deeply and inextricably situated in specific socio-cultural contexts. The authors continuously adapt their language, vocabulary, and rhetorical strategies to address the specific needs, controversies, and cultural backgrounds of their respective audiences—whether addressing the philosophical Greeks in Corinth, the law-focused Jewish Christians in Rome, or the spiritually confused believers in Galatia. This register variation highlights the immense functional flexibility of New Testament Greek. While the Gospels utilize a more straightforward, sequential narrative register, the epistles employ a highly sophisticated, logical, and argumentative register designed explicitly to persuade, instruct, correct, and maintain social bonds within the fragile early Christian communities.

5.3. Ethnographic Analysis of Sacred Dialogue

Dialogic sequences embedded within biblical narratives present a unique and fascinating text type that requires specialized analytical tools. These are not merely transcripts of conversations; they are highly crafted literary and theological constructs. The integration of the ethnography of communication with text linguistics offers a uniquely powerful method for unpacking these interactions.

The recent, groundbreaking study by Agbeleoba et al. (2026) on the brief but profound conversation between Jesus and the two thieves on the cross serves as a perfect paradigm for this methodological approach. By applying Dell Hymes's comprehensive SPEAKING model to this specific New Testament dialogue, Agbeleoba et al. (2026:61) systematically analyzed every facet of the communicative event:

- Setting: The physical and psychological extremity of the crucifixion at Golgotha.
- Participants: Jesus (the central figure of authority, despite his physical state), the repentant thief, and the impenitent thief.
- Ends: The radically different goals of the participants—mockery and a demand for physical rescue versus a humble plea for spiritual salvation.
- Act sequences: The precise order and structure of the verbal exchange.
- Key: The shifting emotional tone of desperation, cynical doubt, nascent faith, and ultimate divine authority.
- Instrumentalities: The spoken language (likely Aramaic, recorded in Greek) under conditions of extreme physical duress.
- Norms: The prevailing socio-religious expectations regarding crime, punishment, divine intervention, and the afterlife.
- Genre: A complex hybrid of sacred dialogue, mockery, and petitionary prayer.

From a strict text linguistic perspective, this dialogue is a highly condensed, intensely intentional text. The impenitent thief's utterance ("Are you not the Christ? Save yourself and us!") functions pragmatically as both a challenge and an insult. It utilizes interrogative and imperative grammatical forms to express profound doubt and a desperate instinct for self-preservation.

In sharp contrast, the repentant thief's discourse demonstrates a complex cognitive, moral, and theological shift. His initial rebuke of the other thief establishes a moral baseline and acknowledges their shared guilt. His subsequent direct plea to Jesus ("Jesus, remember me when you come into your kingdom") is grammatically imperative, but pragmatically, it functions not as a demand, but as a deeply humble petition and an astonishing confession of faith in a dying man's future kingship (Agbeleoba et al., 2026:63).

Jesus's response ("Truly, I say to you, today you will be with me in paradise") serves as the communicative and theological climax of the entire event. The use of the solemn, formulaic introduction "Truly, I say to you" (amen lego soi) immediately establishes a register of absolute, transcendent divine authority. Furthermore, the specific temporal marker "today" creates immediate textual and theological coherence; it directly and shockingly answers the thief's future-oriented, uncertain plea ("when you come") with an assertion of immediate, present reality and assurance.

This combined ethnographic and textual analysis conclusively demonstrates that biblical dialogue is never merely reported speech. It is a carefully constructed, highly stylized text type where specific linguistic choices—the use of pronouns, verb moods, and temporal markers—intersect seamlessly with deep socio-cultural norms to convey profound, multi-layered theological meaning.

6. Discussion

The systematic application of corpus linguistics and multi-dimensional text typology to the Old and New Testaments yields several highly significant insights that advance both biblical scholarship and the broader field of linguistics.

First and foremost, this methodology provides rigorous, empirical validation for the intuitive genre classifications that have long dominated traditional biblical studies. By precisely quantifying linguistic features, researchers are no longer reliant solely on subjective interpretation; they can objectively demonstrate, with statistical backing, exactly how poetic texts differ structurally from legal codes, or how the linguistic profile of epistolary discourse contrasts with that of apocalyptic literature. This solid empirical foundation significantly strengthens the analytical rigor and defensibility of biblical exegesis.

Second, a corpus-based approach brilliantly highlights the internal linguistic diversity and complexity of individual biblical books. Recognizing that a single text, such as the Book of Isaiah or the Book of Exodus, is not a monolith but contains multiple, distinct linguistic text types allows for a much more nuanced and accurate interpretation. Exegetes and translators can better understand and convey how biblical authors masterfully shifted registers and altered their syntax to achieve specific communicative purposes—whether they were narrating history, delivering divine oracles, or prescribing intricate covenantal laws.

Third, the seamless integration of discourse analysis and the ethnography of communication—as vividly demonstrated by the analysis of dialogic texts (Agbeleoba et al., 2026:65)—reveals the deep pragmatic and socio-cultural dimensions of biblical language. It underscores the vital concept that the Bible is not merely a static repository of abstract theological propositions. Rather, it is a dynamic record of communicative events where language functions actively as a medium for negotiating identity, expressing belief, challenging power structures, and performing social action.

Finally, the study of biblical text typology contributes immense value to the broader, secular field of linguistics. The Bible provides a massive, highly structured, and historically rich corpus that is ideal for analyzing long-term language variation, tracking register shifts, and studying the evolution of complex discourse structures over millennia. The transition from the ancient Hebrew and Aramaic of the Old Testament to the Hellenistic Greek of the New Testament offers an unparalleled, unique opportunity for linguists to study how specific text types and rhetorical strategies are translated, adapted, and transformed across vast linguistic, cultural, and temporal boundaries.

7. Conclusion

This article has presented a comprehensive, rigorous, and corpus-based text linguistic exploration of the Old and New Testaments. By systematically applying the principles of text typology, multi-dimensional analysis, systemic functional linguistics, and the ethnography of communication, the study has conclusively demonstrated that the biblical corpus is a highly complex, brilliantly constructed, and multi-layered linguistic ecosystem. The integration of empirical, quantitative corpus linguistics with qualitative functional linguistics provides a highly robust, defensible framework for identifying, categorizing, and deeply interpreting the incredibly diverse text types that comprise the sacred texts.

The quantitative analysis of the vast biblical corpus reveals distinct, empirically verifiable linguistic profiles for all major biblical genres. This highlights exactly how variations in syntax, lexical density, pronoun usage, and verb tense correlate directly with specific communicative functions and authorial intents. Furthermore, the qualitative exploration of narrative cohesion in the Old Testament, the persuasive pragmatics of New Testament epistles, and the ethnographic analysis of sacred dialogue—particularly the profound examination of the conversation between Jesus and the thieves on the cross (Agbeleoba et al., 2026:66)—vividly illustrates how textual coherence and deep socio-cultural context intersect to create enduring theological meaning.

The rigorous application of modern linguistic methodologies to these ancient texts successfully bridges the gap between empirical language study and theological exegesis. A corpus-based approach to biblical text typology does not diminish the sacred nature of the text; rather, it refines our understanding of exactly how these texts were constructed and enriches our appreciation of their profound, enduring communicative power. As computational tools become more sophisticated and corpus resources continue to expand and evolve, the dynamic intersection of text linguistics and biblical studies promises to yield even deeper, more transformative insights into the magnificent linguistic architecture of the Old and New Testaments.

References

Agbeleoba, S. O., Dada, S. A., Afolayan, S. O., & Agbeleoba, P. A. (2026). Ethnographic analysis of 'SPEAKING' and textuality in the conversation between Jesus Christ and the thieves on the Cross. Advances in Social Sciences and Management, 3(5), 56–66. https://doi.org/10.63002/assm.305.1091

Berlin, A. (1989). Lexical cohesion and biblical interpretation. Hebrew Studies, 30, 22–30.

Biber, D. (1989). A Typology of English texts. Linguistics, 27(1), 3–43. https://doi.org/10.1515/ling.1989.27.1.3

Halliday, M. A. K., & Hasan, R. (1989). Language, context, and text: Aspects of language in a social-semiotic perspective. Oxford University Press.

O'Donnell, M. B. (2005). Corpus linguistics and the Greek of the new testament. Sheffield Phoenix Press.

Toffelmire, C. M. (2016). A discourse and register analysis of the prophetic book of Joel (Studia Semitica Neerlandica, 66). Brill.

Walton, J. H. (2019). Old testament theology for Christians: From ancient context to enduring belief. IVP Academic.

Yount, C. (2021). Linguistic dating of biblical texts. Bible Interp. University of Arizona.