

Semantic Coherence in AI-Generated and Human-authored Short Stories

Norah U. Okolie

Department of English and Literary Studies, Enugu State University of Science and Technology, Agbani

Abstract

This paper examines semantic coherence in AI-generated short stories in comparison with human-authored short stories. Drawing on two thematically related prompts derived from award-winning Commonwealth short stories, comparable narratives were generated using ChatGPT and analysed alongside the original human-authored texts. Using a qualitative research design, the study adopts Halliday and Hasan's (1976) theory of cohesion and Van Dijk's coherence principles as its theoretical framework. The analysis presents how linguistic elements are deployed to create logical connectivity, continuity of meaning, and overall textual unity, while also identifying points of incoherence in the AI-generated narratives. Findings show that although the AI-generated stories demonstrate substantial use of cohesive devices, such as reference, conjunction, ellipsis, and lexical repetition, they exhibit weaknesses in both local and global coherence. These weaknesses manifest particularly in illogical action sequencing, culturally implausible scenarios, and insufficiently motivated narrative transitions. The study concludes that while AI can approximate surface-level coherence through linguistic cohesion, deeper semantic and pragmatic coherence, especially within culturally grounded narratives, remains more consistently realised in human-authored texts. By foregrounding the distinction between cohesion and coherence in culturally situated narratives, the study contributes a discourse-based framework for evaluating AI-generated creative texts beyond surface fluency.

Keywords: Semantic coherence, AI-generated narratives, Discourse analysis, Cohesion and coherence, Human–AI text comparison

Introduction

The interpretation of the meaning of a sentence is determined by its relationship with other sentences within a discourse. This relationship accounts for the level of coherence a discourse achieves, as coherence is a fundamental semantic property of discourse (Van Dijk, 1977). Language does not function as a disconnected or isolated set of unrelated sentences; rather, it operates as a collocated, structured, and coherent sequence of sentences that together constitute discourse. It is this relational structure that distinguishes a meaningful discourse from a random assemblage of sentences (Jurafsky & Martin, 2019).

Van Dijk (1977) categorises coherence into two broad types: local (linear or sequential) coherence and global coherence. Local coherence refers to the coherence relations holding between propositions expressed in adjacent or sequential sentences, whereas global coherence characterises the overall unity and interpretability of a discourse or larger discourse fragment (Toolan, 2013). A discourse is considered coherent when ideas progress logically, events are meaningfully connected, and propositions align with shared knowledge and expectations.

In discussing the relationship between text and discourse, Halliday and Hasan (1976) argue that a text is adjudged coherent when it exhibits cohesive ties and aligns with the context in which

it is produced. These cohesive ties contribute to what they term the *texture* and *textual unity* of a text. Hasan (1984a) further maintains that the degree to which cohesive elements are deployed in a text significantly determines its level of coherence. However, cohesion alone does not automatically guarantee coherence. Flowerdew (2013, as cited in Tamunoobelem, 2018) emphasises that there is a graduation in the extent to which cohesion and cohesive harmony contribute to coherence, suggesting that coherence operates beyond surface linguistic connections.

Toolan (2013) conceptualises coherence as encompassing clarity, orderliness, logicity, reasonableness, and overall sense-making, noting that it has traditionally been treated as a text-linguistic notion. However, he argues that narrative coherence extends into the domain of pragmatics because it requires sensitivity to culturally situated meanings and context-specific interpretation. From a text-linguistic perspective, Toolan identifies subtypes of coherence such as temporal coherence, causal coherence, thematic coherence, and topic maintenance and development. Narratives, he argues, are expected to demonstrate unity, continuity, and perseverance of topic, often achieved through repetition and near repetition, including lexical repetition and semantic recurrence.

Chatman (1978, as cited in Toolan, 2013) similarly notes that narrative perseverance is achieved through the repeated naming of characters and events, which functions as a coherence strategy. For instance, a sequence in which a character has an accident, later dies, and is eventually buried demonstrates a string of continuity that sustains narrative coherence. These strings of continuity correspond to what Halliday and Hasan (1976) classify as grammatical and lexical cohesive devices, including reference (e.g., pronouns and demonstratives), substitution, ellipsis, conjunction, and lexical cohesion such as repetition and synonymy.

Despite this close relationship, a text may be cohesive without being coherent. Tamunoobelem (2018) argues that coherence ultimately depends on a text's ability to connect with the reader's familiar schemata of ideational and interpersonal frames of reference. In other words, coherence is achieved when a text aligns with the socio-cultural realities, expectations, and contextual knowledge of its intended audience. This distinction between cohesion and coherence is particularly crucial when evaluating texts produced outside human experiential contexts, such as those generated by artificial intelligence.

Artificial Intelligence and Text Generation

Artificial Intelligence (AI) has recently witnessed rapid advancement and growing scholarly attention, particularly in the domain of natural language generation. Large Language Models (LLMs), such as ChatGPT, have demonstrated impressive capabilities in generating texts across a wide range of communicative tasks, including document writing, question answering, and email composition (Wang et al., 2024). These developments have led to increasing use of AI-generated texts in both academic and non-academic contexts.

However, alongside these advancements, scholars have raised serious concerns regarding the potential misuse of AI-generated texts, including misinformation, plagiarism, and deceptive authorship practices (Adelani et al., 2020; Stiff & Johansson, 2022; Wang et al., 2024; Zellers et al., 2019). In response, a growing body of research has focused on developing methods for detecting AI-generated texts through both internal and external linguistic evidence (Wang et al., 2024). Tang et al. (2023, as cited in Wang et al., 2024) observe that AI-text detection is often

framed as a binary classification problem, primarily aimed at distinguishing AI-generated texts from human-authored ones based on linguistic features.

While AI-generated texts often exhibit high levels of grammatical accuracy and surface fluency, scholars continue to report persistent challenges related to global discourse coherence, including abrupt topic shifts, logical inconsistencies, and contradictory propositions (Kotonya & Toni, 2020). These challenges suggest that although AI systems are proficient at producing cohesive sentence sequences, their ability to sustain semantic, pragmatic, and culturally grounded coherence across extended discourse remains uncertain.

Despite growing interest in AI-generated discourse, much existing research prioritises detection mechanisms, automated metrics, or surface-level linguistic features, with relatively limited attention to semantic coherence as a discourse-level phenomenon. Moreover, few studies conduct qualitative, theory-driven comparisons between AI-generated and human-authored narratives using established discourse frameworks. This gap is particularly evident in analyses of short stories, where narrative coherence relies heavily on logical sequencing, motivation, and cultural plausibility.

Addressing this gap is crucial for advancing understanding of the limits and capabilities of AI in narrative meaning-making. A discourse-based examination of cohesion and coherence offers deeper insights into how AI-generated narratives function beyond grammatical correctness and stylistic fluency. This study aims to examine and compare the semantic coherence of AI-generated and human-authored short stories by analysing how cohesive devices contribute to, or fail to achieve, coherent narrative meaning. Specifically, the study seeks to identify and analyse the cohesive devices deployed in both AI-generated and human-authored short stories, drawing on Halliday and Hasan's (1976) cohesion framework; evaluate how these cohesive devices contribute to semantic coherence in the narratives, using Van Dijk's coherence principles as an interpretive guide; and examine and explain instances of incoherence in AI-generated short stories, with particular attention to logical sequencing, narrative motivation, and cultural plausibility.

This study is guided by the following research questions:

1. What cohesive devices characterise AI-generated and human-authored short stories?
2. How do these cohesive devices contribute to local and global semantic coherence in the narratives?
3. In what ways do AI-generated short stories exhibit incoherence when evaluated against narrative logic and cultural plausibility?

Literature Review

Artificial Intelligence and Narrative Text Generation

Recent advances in artificial intelligence, particularly large language models (LLMs), have significantly transformed the landscape of automated text generation. Models such as GPT-3, GPT-4, and subsequent iterations have demonstrated remarkable capacity to produce grammatically accurate, stylistically fluent, and contextually appropriate texts across a wide range of genres, including storytelling, journalism, and academic writing. As a result, scholarly attention has increasingly shifted from surface-level accuracy to deeper questions concerning meaning construction, discourse organisation, and narrative coherence in AI-generated texts.

Early studies on AI-generated narratives largely focused on syntactic correctness and lexical diversity, often concluding that machine-generated texts approximate human writing at the sentence level (Brown et al., 2020; Floridi & Chiriatti, 2020). However, subsequent research has problematised these findings by arguing that grammatical well-formedness does not necessarily equate to coherent discourse. Scholars such as Marcus and Davis (2019) and Bender et al. (2021) caution that language models operate primarily through probabilistic pattern matching rather than genuine semantic understanding, which may result in texts that appear fluent but lack deeper narrative logic.

Within narrative studies, coherence is particularly critical because storytelling relies not only on sentence-level cohesion but also on causal progression, character motivation, and culturally grounded plausibility. While AI systems can replicate narrative templates, several studies indicate that they struggle to maintain consistent plot development and meaningful thematic continuity over extended discourse (Kreminski et al., 2022; Peng et al., 2023). These limitations become more pronounced in culturally specific narratives, where shared social knowledge and pragmatic norms play a central role in meaning-making.

Cohesion and Coherence in Discourse Analysis

The distinction between cohesion and coherence has long occupied a central position in discourse analysis. Halliday and Hasan's (1976) seminal work defines cohesion as the network of linguistic ties, such as reference, substitution, ellipsis, conjunction, and lexical cohesion, that connect elements within a text. From this perspective, cohesion is observable at the textual surface and can be identified through explicit linguistic markers.

Coherence, by contrast, operates at a deeper semantic and pragmatic level. Van Dijk (1977, 1980) conceptualises coherence as the logical and conceptual unity of a discourse, emphasising the role of macrostructures, world knowledge, and inferencing processes in text interpretation. Unlike cohesion, coherence is not always explicitly encoded in linguistic forms but is constructed through readers' interpretation of how ideas relate meaningfully within a given context.

Subsequent scholarship has reinforced the view that cohesion is neither a sufficient nor a guaranteed condition for coherence. Texts may be highly cohesive yet semantically incoherent if ideas are poorly motivated or logically inconsistent (Enkvist, 1990; Widdowson, 2007). Conversely, texts with minimal overt cohesive markers may still achieve coherence through shared cultural schemas and contextual inference. This distinction is especially relevant in the evaluation of AI-generated texts, which often display abundant cohesive devices without necessarily achieving conceptual unity.

Coherence in AI-Generated Texts

Recent empirical studies examining AI-generated discourse reveal a recurring pattern: strong performance in cohesion alongside weaknesses in coherence. For instance, studies analysing AI-generated essays and narratives report frequent use of referential ties, conjunctions, and lexical repetition, suggesting that language models are adept at reproducing surface-level textual connectedness (Guan et al., 2022; Zellers et al., 2019). However, closer qualitative analyses indicate that these cohesive features do not always support logical progression or meaningful discourse development.

In narrative contexts, AI-generated texts have been shown to exhibit abrupt topic shifts, inconsistent character actions, and implausible event sequencing (Karinshak et al., 2021). These issues are often attributed to the absence of long-term discourse planning and limited access to real-world experiential knowledge. While AI models can statistically predict plausible next sentences, they lack an embodied understanding of social norms, cultural practices, and narrative causality.

Furthermore, research suggests that coherence breakdowns become more visible when AI-generated texts are evaluated against culturally grounded narratives rather than generic prompts (Ji et al., 2023). Cultural narratives require adherence to shared practices, values, and symbolic meanings that are not always explicitly stated but are assumed within the discourse community. When such assumptions are violated, texts may appear incoherent or implausible to human readers, even if they remain grammatically sound.

Human vs AI Writing: Comparative Perspectives

Comparative studies between human-authored and AI-generated texts consistently highlight differences in semantic depth, pragmatic sensitivity, and discourse intentionality. Human writers draw on lived experience, cultural knowledge, and communicative intent, enabling them to construct narratives that align with shared social realities. AI systems, in contrast, rely on statistical correlations derived from training data, which may not adequately capture culturally specific meaning systems.

Research comparing AI and human storytelling indicates that while AI can emulate narrative structure at a formal level, it struggles with motivation-driven actions and culturally appropriate behaviours (Clark et al., 2021). Human-authored narratives tend to demonstrate stronger global coherence, characterised by consistent thematic development and logically connected events. AI-generated narratives, on the other hand, often prioritise local coherence at the sentence or paragraph level, sometimes at the expense of overall narrative unity.

Importantly, scholars argue that evaluations of AI writing should move beyond automated metrics and incorporate qualitative, discourse-based analyses that foreground meaning, context, and interpretation (Bender et al., 2021; Swales, 2020). Such approaches allow researchers to uncover subtle incoherencies that may be overlooked by surface-level assessments.

Identified Research Gap

Despite growing scholarly interest in AI-generated discourse, there remains a notable gap in studies that systematically compare semantic coherence in AI-generated and human-authored narratives using established discourse theories. Much of the existing research either prioritises technical performance metrics or focuses narrowly on cohesion without adequately addressing coherence as a culturally and pragmatically situated phenomenon.

Moreover, relatively few studies employ thematic literary texts as prompts or analyse AI-generated narratives within culturally specific frameworks. This limits understanding of how AI handles narrative plausibility, social norms, and culturally embedded meaning structures. There is therefore a need for qualitative, theory-driven analyses that examine not only whether AI-generated texts are cohesive, but whether they are meaningfully coherent when evaluated against human-authored narratives rooted in specific cultural contexts.

This study addresses this gap by adopting Halliday and Hasan's theory of cohesion alongside Van Dijk's coherence framework to conduct a comparative analysis of AI-generated and human-authored short stories. By focusing on semantic coherence and cultural plausibility, the study contributes to ongoing debates about the capabilities and limitations of AI in narrative meaning-making.

Theoretical Framework

This research is grounded on two frameworks, Halliday and Hassan's (1976) theory of cohesion and Van Dijk's theory of coherence. Cohesion, according to Halliday and Hassan, refers to the visible linguistic elements that connect to each other to make a meaningful whole structure. There are grammatical and lexical relationships that link different parts of a text. The theory posits that these cohesive devices exist to unify and bring about a coherent text. Their model identifies five types of cohesive devices namely, reference made up of pronouns, demonstratives, and comparatives that refer to other elements within the text; substitution meaning the replacement of a word or phrase with another; ellipsis indicating the omission of a word or phrase; conjunction linking words and phrases (e.g., and, but, however); and lexical cohesion which detects repetition, synonymy, hyponymy, and collocations which are drawn from specific semantic fields. This theory provides a framework to decipher and measure the semantic coherence of a text through the lexical relations in the text.

Van Dijk, in his theory of coherence, categorizes the coherence of any discourse or text into local and global coherence. Local coherence deals with how sentences are connected to each other in transmitting meaning; how a sentence close to another sentence in a discourse contains lexical elements that can be meaningfully linked to the previous sentence; and the sentence-to-sentence flow of a discourse. For example, the use of a pronoun 'she' after using the noun phrase 'the woman' in these two clauses: 'The woman shouted at her husband. She apologized later.' genuinely links the two lexical items as one replaces the other, but the same referent is meant. Also, the relationship between the two clauses is revealed through the cause-effect narration in the text, giving the reason why the woman apologized. Global coherence, on the other hand, explains when a text is united as a whole or not. In global coherence, the various parts that are supposed to focus on the main theme or goal of the text are evaluated. Global coherence depends on the macro-structures, genre and what a reader makes of a text. For example, if the theme of a discourse focuses on cooking, then terms like 'cooking utensils', 'ingredients' and the various processes of the particular thing that is being cooked are expected. If the genre is that of an academic article, some sections like Methodology and Results are expected. The relevance and the logical progression of what is being said are well considered in global coherence.

Methodology

This study adopts a qualitative research design to examine and compare the semantic coherence of AI-generated and human-authored short stories. A qualitative approach is considered appropriate because coherence and cohesion are discourse-level phenomena that require close textual interpretation rather than statistical generalisation.

Two award-winning short stories were purposively selected as the human-authored data. These texts were written by Akwaeke Emezi and Innocent Chizaram Ilo, both winners of the

Commonwealth Short Story Prize. The stories were chosen because they are grounded in Igbo cultural contexts, with which the researchers are familiar. This cultural familiarity is crucial for assessing global coherence, narrative motivation, and cultural plausibility in line with Van Dijk's coherence framework.

To enable systematic comparison, two AI-generated short stories were produced using ChatGPT. The AI texts were generated by supplying carefully designed prompts that mirrored the storyline structure, themes, cultural setting, and approximate length of the human-authored short stories. This prompt-guided generation ensured comparability between the datasets and reduced variability arising from thematic or stylistic mismatch.

The short story by Akwaeke Emezi contains 4,123 words, while that of Innocent Chizaram Ilo contains 3,231 words, in line with the word-length requirements of the Commonwealth Short Story Prize. For analytical manageability and consistency, the scope of analysis was delimited to the first 1,500 words of each text, both human-authored and AI-generated. This delimitation was adopted to allow detailed and systematic analysis of grammatical and lexical cohesive devices, as well as local and global coherence features, without compromising analytical depth.

The AI-generated texts analysed in this study were produced solely for research purposes using controlled prompts designed to mirror the structure and themes of the selected human-authored short stories. No human-authored text was input into the AI system beyond thematic and structural descriptions. The study does not attribute authorship or creative agency to the AI-generated narratives, and all analyses treat them strictly as machine-produced linguistic artefacts.

Data analysis proceeded in two complementary stages. First, Halliday and Hasan's (1976) cohesion framework was employed to identify and analyse grammatical and lexical cohesive devices in both the AI-generated and human-authored short stories. These included reference, substitution, ellipsis, conjunction, and lexical cohesion (such as repetition, synonymy, and semantic recurrence).

Second, Van Dijk's theory of coherence was applied to evaluate how these cohesive devices contribute to semantic coherence at both the local and global levels. Particular attention was paid to logical sequencing, topic continuity, narrative progression, and the consistency of propositions across sentences and larger discourse units. Instances where cohesive ties failed to produce coherent meaning, especially in the AI-generated texts, were examined and interpreted in relation to narrative logic, pragmatic expectations, and cultural plausibility.

Through this combined analytical procedure, the study systematically compares AI-generated and human-authored short stories to determine how cohesion operates as a resource for, or a limitation to, achieving coherent narrative discourse.

Presentation and Analysis of Data

Analysis of AI-Generated Short Stories

Reference

All three types of reference, personal, demonstrative, and comparative, were employed in the AI-generated short stories *Powder on His Skin* (Story 1) and *Breasts of the Earth* (Story 2). Personal reference was the most frequently used cohesive device and played a central role in character continuity and narrative progression.

In *Powder on His Skin*, for instance, the personal possessive pronoun *his* appears twice in the opening paragraph to link the protagonist, Chibuike, to *his chest* and *his sister's humming*. Without the first *his*, the referent *chest* would lose specificity, thereby disrupting semantic coherence. Similarly, the second *his* establishes the familial relationship between Chibuike and his sister, Nneka, embedding emotional memory within the narrative. Across subsequent paragraphs, personal pronouns such as *he*, *she*, and *they* consistently maintain referential clarity: *he* refers to Chibuike, *she* to Nneka or the mother depending on context, and *they* to village children who mock Chibuike.

A comparable pattern is evident in *Breasts of the Earth*, where *she* predominantly refers to Nwanyibuife, *he* to her husband Uchechukwu, and *they* to the group of women. These sustained referential links enhance continuity and prevent referential ambiguity, thereby supporting local coherence.

Demonstrative references (*this*, *that*, *these*, *those*) were employed to anchor specific experiences and narrative moments, as seen in expressions such as “*this abomination came from you*” and “*that was when he heard the gasp*” (*Powder on His Skin*), and “*that night her husband beat her*” and “*this refusal...*” (*Breasts of the Earth*). Comparative references (*like*, *as though*) were also used to create imagery linked to familiar experiential frames, for example, “*like birds warming the morning*” and “*like a rope round her throat*”.

Taken together, these reference devices function as cohesive ties that enhance continuity and narrative flow. However, as Van Dijk (1977) argues, cohesion alone does not guarantee coherence, a distinction that becomes evident in later sections.

Substitution and Ellipsis

Substitution occurred sparingly in the AI-generated texts. A notable example is the nominal substitution “*This one is soft*”, where *this* substitutes for Chibuike and encodes evaluative judgement from his father. While cohesive, such substitutions were limited in frequency.

Ellipsis, however, produced mixed effects. Some elliptical constructions were pragmatically interpretable, such as the dialogue exchange “*They have finished?*” – “*Yes.*”, which mirrors natural spoken interaction. In contrast, other elliptical expressions obscured meaning and disrupted coherence. The sentence “*Sometimes she let him touch.*”, presented as a standalone line, lacks a recoverable referent, leaving it unclear whether the object is dough, a hand, or something else. This ambiguity undermines both local and global coherence by interrupting logical progression.

Conjunctions

Additive conjunctions, particularly *and*, dominated the AI-generated texts. They were used to accumulate actions and descriptions, for example, “*She rushed forward and slapped him.*” Adversative conjunctions such as *but* introduced contrast (“*She still loved him, but her world had widened*”), while causal (*so*, *because*) and temporal (*when*, *then*, *before*, *after*) conjunctions explained motivation and sequencing.

Although these conjunctions establish surface-level connectivity, their over-reliance, especially on additive forms, sometimes results in linear accumulation rather than logical

development, echoing concerns raised by Kotonya and Toni (2020) regarding AI-generated discourse progression.

Lexical Cohesion: Repetition, Relations, and Collocation

Lexical repetition was used extensively to maintain thematic focus. Names such as *Chibuike*, *Mama*, and *Nneka* in the first story, and *Nwanyibuife* and *Uchechukwu* in the second, function as perseverative devices (Chatman, 1978; Toolan, 2013). Superordinate relations were also evident, for instance, *makeup* subsuming *powder*, *brush*, *lipstick*, and culinary items such as *egg*, *sugar*, *dough* forming a baking semantic field.

Collocations further strengthened thematic clustering. In *Powder on His Skin*, lexical items such as *kneading*, *flour*, *eggs*, *sugar*, *bowl*, and *kitchen* construct a culinary semantic field, while *slap*, *scream*, *punishment*, and *drag* signal conflict. In *Breasts of the Earth*, collocations such as *breasts*, *womb*, and *children* foreground fertility and womanhood.

These lexical patterns enhance textual cohesion but, as demonstrated below, do not consistently translate into coherent narrative logic.

Local and Global Coherence Issues in AI-Generated Stories

Despite cohesive density, several narrative scenarios in the AI-generated texts raise coherence concerns when evaluated against cultural plausibility and logical sequencing.

Powder on His Skin

First, the depiction of bread frying within an Igbo cultural setting is contextually questionable. Bread is typically baked rather than fried in Nigerian-Igbo contexts, and the absence of explanatory context undermines local coherence. Similarly, the claim that a seven-year-old girl routinely ties a slipping shirt button raises questions of feasibility, given that such an action would normally require sewing tools and instruction.

Additionally, the repeated hand-squeezing gesture attributed to a seven-year-old lacks narrative motivation or explanation, weakening plausibility. As Van Dijk (1977) notes, local incoherence accumulates to affect global discourse coherence.

Breasts of the Earth

The progression of events surrounding Nwanyibuife's beating by her husband lacks logical sequencing. While the dialogue suggests a calm refusal, the sudden escalation to physical violence is not narratively motivated. Spatial relations, such as the distance between characters, are unspecified, making the described actions difficult to reconcile.

Further incoherence arises in the portrayal of the *umuada*. The group is depicted engaging in domestic chores such as plaiting hair and shelling melon seeds, activities that conflict with the culturally recognised institutional role of the *umuada* in Igbo society. This misrepresentation disrupts global coherence by violating shared cultural schemata (Tamunoobelem, 2018).

Table 1: Summary of Cohesion and Coherence Features in AI-Generated Stories

Feature	Observation	Effect
Reference	High use of personal pronouns	Strong cohesion

Ellipsis	Some unclear omissions	Coherence breakdown
Conjunctions	Overuse of additive <i>and</i>	Linear accumulation
Lexical cohesion	Strong thematic clustering	Surface coherence
Cultural plausibility	Weak in key scenes	Global incoherence

Analysis of Human-Authored Short Stories

The human-authored stories exhibit similar cohesive devices but differ significantly in coherence outcomes.

Reference, Substitution, and Ellipsis

Personal references (*I, my, she, her, we*) dominate the narratives and are used with high referential clarity. Ambiguous references, such as *you* in *When a Woman Renounces Motherhood*, function pragmatically to address multiple addressees, enriching rather than obscuring meaning.

Substitution and ellipsis occur naturally and reflect authentic spoken interaction, for example, “*Not bad*” standing in for “*This cake is not bad*”. Unlike the AI texts, these omissions are recoverable and contextually motivated.

Lexical Networks and Semantic Fields

Human-authored texts demonstrate dense and culturally grounded semantic networks. Baking-related lexis (*mix, stir, icing*) supports thematic unity, while emotion-related vocabulary (*fear, guilt, love*) develops psychological depth. Cultural lexical items (*nzu, lappah, Oriego Age Group*) are accurately embedded, reinforcing global coherence.

Table 2: Comparative Summary of AI-Generated and Human-Authored Narratives

Dimension	AI-Generated Stories	Human-Authored Stories
Cohesive density	High	High
Local coherence	Generally sustained	Consistently sustained
Global coherence	Frequently unstable	Strong and culturally grounded
Cultural plausibility	Occasionally violated	Consistently maintained
Narrative motivation	Often underdeveloped	Psychologically and socially motivated

Discussion of Findings

This study set out to examine and compare the semantic coherence of AI-generated and human-authored short stories through a qualitative analysis of cohesive devices and their contribution to narrative coherence. The findings reveal a clear distinction between surface-level cohesion and deep-level semantic coherence, particularly in culturally situated narratives.

Consistent with Halliday and Hasan’s (1976) framework, both AI-generated and human-authored texts demonstrate extensive use of grammatical and lexical cohesive devices, including reference, conjunction, ellipsis, repetition, and collocation. In the AI-generated stories, the density of these cohesive ties, especially personal reference and lexical repetition, creates an initial impression of textual connectedness. Characters are consistently named and pronominally tracked,

events are sequentially linked through conjunctions, and thematic clusters are maintained through repeated lexical fields such as domestic labour, violence, and family relations. At the level of cohesion, therefore, the AI texts appear structurally competent and stylistically fluent.

However, when evaluated through Van Dijk's (1977) distinction between local and global coherence, these cohesive patterns do not consistently translate into meaningful narrative logic. Several instances in the AI-generated stories reveal breakdowns in causal motivation, cultural plausibility, and inferential continuity, particularly where actions or events are insufficiently motivated or contradict shared socio-cultural knowledge. This supports the position advanced by Toolan (2013) that coherence extends beyond formal textual links to include pragmatic and cultural interpretability.

A central finding of this study is that AI-generated narratives tend to rely on linear accumulation of cohesive elements rather than hierarchical organisation of narrative meaning. The frequent use of additive conjunctions, for example, produces a chaining effect in which events are placed side by side without adequate causal grounding. While such chaining maintains local continuity, it often fails to establish why events occur, how characters are psychologically positioned, or how actions align with culturally recognised practices. This observation aligns with Kotonya and Toni's (2020) claim that AI-generated texts may exhibit fluent sentence-level transitions while struggling with discourse-level reasoning and narrative planning.

The analysis further demonstrates that ellipsis functions differently across the two datasets. In the human-authored stories, elliptical constructions are largely recoverable and pragmatically motivated, reflecting naturally occurring spoken and written discourse. In contrast, some elliptical expressions in the AI-generated texts obscure referential clarity, resulting in interpretive gaps that disrupt coherence. This supports Flowerdew's (2013) assertion that cohesion contributes to coherence in graduated degrees, rather than in an automatic or uniform manner.

One of the most significant divergences between the AI-generated and human-authored narratives lies in the handling of cultural schemata. Human-authored stories draw on culturally embedded knowledge, such as domestic practices, kinship roles, religious beliefs, and indigenous institutions, to structure narrative progression and reader inference. These shared schemata allow readers to fill gaps, interpret motivations, and anticipate outcomes, thereby strengthening global coherence. The AI-generated stories, by contrast, occasionally misrepresent or trivialise culturally specific institutions, such as the *umuada*, or depict actions that are contextually improbable within the Igbo socio-cultural environment. These misalignments weaken coherence not because of linguistic error, but because of pragmatic and cultural incongruity, echoing Tamunoobelem's (2018) view that coherence depends on alignment with readers' socio-cultural frames of reference.

Importantly, the findings also suggest that AI's limitations are not primarily grammatical or lexical, but conceptual and pragmatic. The AI-generated texts successfully replicate the formal markers of narrative cohesion but struggle to sustain narrative logic across extended discourse. This distinction reinforces Halliday and Hasan's (1976) claim that cohesion is a necessary but insufficient condition for coherence and extends it to the evaluation of AI-generated creative texts.

In relation to existing AI-text detection research, which has largely focused on statistical classification and surface linguistic features (Zellers et al., 2019; Wang et al., 2024), this study contributes a discourse-analytic perspective that foregrounds meaning-making, cultural grounding, and narrative reasoning. By demonstrating that AI-generated narratives can appear cohesive yet

fail to achieve stable semantic coherence, the study highlights the limitations of approaches that equate fluency with discourse competence.

Conclusion and Recommendations

This study examined the semantic coherence of AI-generated and human-authored short stories through a discourse-analytic lens, drawing on Halliday and Hasan's cohesion theory and Van Dijk's coherence framework. The findings demonstrate that while AI-generated narratives exhibit extensive use of grammatical and lexical cohesive devices, such formal cohesion does not consistently yield coherent meaning at the discourse level. In contrast, human-authored stories achieve stronger global coherence by embedding cohesive ties within culturally grounded narrative logic, inferential continuity, and pragmatically motivated event sequencing.

The study establishes that AI's current narrative competence is largely structural rather than conceptual. Although AI systems can replicate surface patterns of textual cohesion, they remain limited in modelling culturally situated reasoning, causal motivation, and narrative intentionality. These limitations are most evident in contexts requiring shared socio-cultural knowledge, where semantic coherence depends not only on linguistic connectedness but also on culturally informed interpretation.

By foregrounding the distinction between cohesion and coherence in AI-generated creative texts, this study contributes to discourse studies and AI linguistics by offering a qualitative, theory-driven evaluation that moves beyond fluency-based assessments. It reinforces the enduring relevance of discourse frameworks in understanding the boundaries of machine-generated language. Despite its insights, this study is limited by its small dataset and focus on a single cultural context. While this allows for depth of analysis, findings may not be generalisable across genres or cultures.

Based on these findings, future research should adopt larger and more diverse datasets across genres and cultural contexts to further examine coherence patterns in AI-generated discourse. Integrating cognitive and pragmatic models alongside cohesion and coherence frameworks may also yield deeper insights into AI narrative reasoning. For pedagogical and creative applications, AI-generated texts should be approached as supportive drafting tools rather than autonomous narrative agents, particularly in culturally sensitive domains. Finally, AI developers are encouraged to incorporate discourse-level and cultural modelling mechanisms to improve semantic coherence in extended narrative generation.

References

- Adelani, D. I., Mai, H., Fang, F., Nguyen, H. H., Yamagishi, J., & Kawai, H. (2020). Generating sentiment-preserving fake online reviews using neural language models and their human and machine-based detection. *Proceedings of the 28th International Conference on Computational Linguistics*, 2001–2013. <https://doi.org/10.18653/v1/2020.coling-main.181>
- Chatman, S. (1978). *Story and discourse: Narrative structure in fiction and film*. Cornell University Press.

- Emezi, A. (2017). *Who is like God*. Commonwealth Short Story Prize. <https://www.commonwealthwriters.org/prize/commonwealth-short-story-prize/shortlisted-stories/who-is-like-god/>
- Flowerdew, J. (2013). Discourse in English language education. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 404–420). Wiley-Blackwell.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- Hasan, R. (1984). Coherence and cohesive harmony. In J. Flood (Ed.), *Understanding reading comprehension* (pp. 181–219). International Reading Association.
- Ilo, I. C. (2020). *The waiting*. Commonwealth Short Story Prize. <https://www.commonwealthwriters.org/prize/commonwealth-short-story-prize/shortlisted-stories/the-waiting/>
- Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed., draft). Pearson. <https://web.stanford.edu/~jurafsky/slp3/>
- Kotonya, N., & Toni, F. (2020). Explainable automated fact-checking for public health claims. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 7740–7754. <https://doi.org/10.18653/v1/2020.emnlp-main.623>
- Stiff, E., & Johansson, F. (2022). Detecting computer-generated disinformation. *International Journal of Data Science and Analytics*, 13(4), 363–374. <https://doi.org/10.1007/s41060-021-00278-8>
- Tamunoobelem, I. (2018). Cohesion and coherence in selected political speeches in Nigeria. *Journal of Language and Linguistics*, 6(2), 45–60.
- Tang, R., Yang, S., Lu, X., & Lin, J. (2023). A benchmark for human–AI text detection. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, 1256–1273. <https://doi.org/10.18653/v1/2023.acl-long.71>
- Toolan, M. (2013). Coherence. In P. Hühn, J. C. Meister, J. Pier, & W. Schmid (Eds.), *The living handbook of narratology*. Hamburg University Press. <https://www.lhn.uni-hamburg.de/node/22.html>
- Van Dijk, T. A. (1977). *Text and context: Explorations in the semantics and pragmatics of discourse*. Longman.
- Wang, Q., Feng, S., Gao, Y., & Li, D. (2024). Detecting AI-generated text: A survey of methods, datasets, and challenges. *ACM Computing Surveys*, 56(2), Article 31. <https://doi.org/10.1145/3626234>
- Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2019). Defending against neural fake news. *Advances in Neural Information Processing Systems*, 32, 9054–9065.