A Study on the Influence of AI-Assisted Writing on the Style of English Academic Discourse

Cheng Huang[*]

Hainan Vocational University of Science and Technology, Haikou, 571100, China
[*]*Corresponding author: hctotoo@outlook.com*

Abstract: *The impact of AI-assisted writing on the style of English academic discourse has become a focus of academic attention. Through multi-layered linguistic analysis, this study systematically explores the stylistic representations of AI-generated texts in terms of lexis-grammar, rhetorical structure, and human-computer interaction. The study finds that, at the lexical level, the texts exhibit a concentration of high-frequency academic vocabulary and a reliance on nominalization structures. At the syntactic level, they demonstrate a tendency towards formulization and a preference for the passive voice. In terms of discourse cohesion, the texts rely on explicit logical markers, yet exhibit limitations in long-range coherence. Regarding rhetorical structure, AI can achieve standard move progression; however, its identification of research gaps remains superficial. In human-computer interaction, stylistic negotiation and disciplinary conventions shape the heterogeneous nature of the blended texts. This study provides a theoretical framework for understanding the discourse style of AI-assisted writing and offers implications for the teaching of academic writing and technological optimization.*

Keywords: *academic discourse style; AI-assisted writing; lexical-grammatical features; rhetorical structure; human-computer interaction*

Introduction

The rapid development of artificial intelligence technology is reshaping the paradigm of academic writing, with AI-assisted writing now widely applied in the production of English academic discourse. However, existing research predominantly focuses on the level of technological application, leaving a systematic exploration of the stylistic features of AI-generated texts and their formative mechanisms notably insufficient. As an important carrier of disciplinary norms and authorial voice, the extent and manner in which academic discourse style is influenced by AI technology urgently requires analysis. Based on theories of linguistics and discourse analysis, this study investigates the representations of AI-generated texts across three dimensions—lexis-grammar, rhetorical structure, and human-computer interaction — to reveal the mechanisms through which AI reshapes academic discourse style. The research not only provides a theoretical framework for understanding AI-assisted writing but also offers significant references for academic writing practice and the optimization of AI tools.

1. The Evolution of Lexical-Grammatical Features in Academic English with AI Assistance

1.1 Patterns of Academic Vocabulary Complexity and Density Distribution

At the level of lexical complexity, algorithm models trained on large-scale corpora tend to invoke an optimized "core academic vocabulary." This selection mechanism maintains the type-token ratio within a stable range, yet the lexical frequency distribution exhibits distinctive characteristics: the recurrence rate of high-frequency academic vocabulary increases significantly, while the use of highly context-dependent specialized terminology appears selective. This distribution pattern reflects the algorithm's pursuit of linguistic standardization but may affect the complete representation of disciplinary terminological systems.

Regarding lexical density, the proportion of content words in AI-generated texts is generally higher than the benchmark set by human writing. This high-density distribution stems from the model's information compression strategy, which aims to maximize propositional information load within limited syntactic space. Nominalization structures, as a key method for achieving high lexical density, are widely employed. Although they enhance the abstractness of the discourse, they may lead to a

reduction in the dynamic quality of action processes and a blurring of semantic participant roles[1].

In terms of lexical collocation patterns, AI texts demonstrate clear characteristics of statistical learning. Their word co-occurrence relationships largely follow high-frequency combinations found in the training data, with relatively limited instances of creative metaphors or unconventional academic expressions. While this collocation preference ensures the standardization of language output, it may constrain the innovative space for academic expression.

1.2 The Formulization and Variation of Syntactic Structures

The formulization of syntactic structures is a prominent feature of AI-generated texts. Quantitative linguistic analysis reveals that the sentence length distribution curve of AI texts is relatively concentrated, with most sentence lengths falling within a comparatively narrow range. This distribution characteristic reflects the internal mechanism of the model's control over syntactic complexity.

At the level of sentence pattern selection, AI texts exhibit specific structural preferences. The frequency of reporting structures and existential sentences is significantly higher than the benchmark set by human writing. Although these patterns conform to academic conventions, their excessive and concentrated use may lead to a monotonous textual rhythm. Regarding clause usage, attributive clauses and object clauses dominate, while adverbial clauses that express complex logical relationships are relatively limited.

The pattern of passive voice usage is particularly noteworthy. AI texts often mechanically adhere to the "objectivity" convention of academic writing, extensively employing passive structures when the agent is unspecified. While this choice conforms to formal norms, it may unnecessarily weaken the action orientation of the discourse. Overall, while the syntactic structures of AI texts ensure linguistic standardization, they may impact the flexibility and expressiveness of academic exposition.

1.3 Automated Implementation of Cohesive Devices

At the level of discourse cohesion, AI-generated texts demonstrate automated implementation mechanisms based on algorithmic logic. The use of logical connectors exhibits systematic preferences, with their frequency and distribution patterns differing from those in human writing. Connectors expressing contrast, cause-and-effect, and progression form the main framework of the cohesive network. However, the application of these connectors sometimes displays formalized characteristics, whereby cohesive components are mechanically inserted even in contexts with insufficient semantic transition intensity[2].

The anaphoric coherence system exhibits characteristics of distance sensitivity. For referential relationships over short distances, the AI system can accurately establish referential links. However, when the referential distance extends and requires spanning multiple propositional units, the probability of referential ambiguity increases significantly. This characteristic reflects the current model's limitations in processing long-distance semantic dependencies.

Regarding lexical cohesion, the use of synonym substitution and hyponymy reveals algorithmic traits. The AI system can establish formal cohesion through repetition and synonym replacement. However, for semantic prosody and collocational restrictions that require deep disciplinary knowledge for accurate mastery, there remains room for improvement in its processing precision. Overall, the coherence of AI-generated discourse is built upon the systematic deployment of explicit cohesive devices, and its cohesion quality directly depends on the algorithm's depth in modeling the semantic networks of academic discourse.

2. Rhetorical Structure and Information Flow in AI-Generated Academic Discourse

2.1 Move Composition and Information Positioning in the Introduction Section

Regarding the move composition of academic introductions, AI-generated texts exhibit features that are both regular and complex. Using Swales' CARS model as an analytical framework, AI demonstrates powerful information integration capabilities within the "Establishing a Territory" move. It can accurately summarize research backgrounds, cite key literature, and establish the significance of the research topic. This ability stems from the deep learning of vast academic literature by large language models, enabling them to recognize and reproduce consensus-based discursive patterns within a

disciplinary field.

In the crucial "Establishing a Niche" move, AI's performance shows specific tendencies. Its method for identifying research gaps primarily relies on the sensitive detection of explicit gap-indicating expressions (such as "lack of," "few studies") in the literature, as well as the presentation of contrasting viewpoints. However, its ability to identify implicit research gaps—such as methodological limitations, theoretical contradictions, or unverified hypotheses, which require deep disciplinary knowledge and creative thinking—remains insufficient. This limitation often results in constructed research gaps that remain superficial, struggling to address the deeper contradictions within disciplinary development.

In the "Occupying the Niche" move, AI can typically formulatically state research objectives, pose research questions, and preview findings. However, its elaboration of research value often employs standardized templates, and its articulation of research innovation tends to focus on extending and supplementing existing research models rather than achieving genuinely groundbreaking academic positioning. This characteristic reflects the structural challenges current natural language processing technology faces in comprehending the nature of academic innovation, and it illustrates the boundaries of algorithmic models in formally imitating academic discursive conventions. It is noteworthy that AI demonstrates high fluency in handling transitions between the three moves, employing appropriate move transition markers to maintain the overall coherence of the introduction section.

2.2 Thematic Progression and the Construction of Information Hierarchies

As a core representation of information structure, thematic progression patterns in AI-generated academic discourse exhibit unique organizational logic and structural characteristics. Data analysis reveals that the linear thematic progression pattern (where the rheme of a preceding sentence becomes the theme of the following sentence) dominates absolutely in AI texts. While this single mode ensures the continuity and traceability of information transmission, its prolonged use leads to a mechanistic argumentative rhythm, lacking necessary variation and emphasis, which impacts the expressiveness and persuasive effect of academic exposition[3].

Regarding the construction of information hierarchies, AI texts demonstrate a notable reliance on explicit markers. The frequent use of guiding phrases such as "it is important to note" and "the key point lies in" to delineate information importance represents a clear strategy. While this approach can explicitly signal information weight, it may substitute for the implicit stratification of information typically achieved through complex syntactic structures and rhetorical devices. Further observation indicates that the distribution of information focus within AI-generated paragraphs is relatively even, lacking the information peaks created by human authors through carefully designed syntactic arrangements and positional effects. This homogenizing treatment may weaken the prominence of key arguments.

Derived thematic progression and hyperthematic progression occur with lower frequency in AI texts. These two progression patterns, which should serve the deepening and transition of arguments, are relatively scarce, potentially resulting in insufficient three-dimensionality in argumentation. Furthermore, when dealing with multi-layered complex arguments, AI often maintains surface coherence by adding cohesive devices rather than reflecting logical layers through organic variation in thematic structure. This difference highlights the current model's limitations in deep logical structuring.

2.3 The Use of Metadiscourse Resources and the Interjection of Authorial Stance

Metadiscourse, as a crucial resource for authors to guide readers and construct academic identity, exhibits systematic distribution features and application patterns in AI-generated academic discourse. The frequency of interactive metadiscourse — including logical connectives, frame markers, and endophoric markers—in AI texts is significantly higher than in human writing. This overuse reflects the algorithm's pursuit of discursive structural clarity. However, dense guiding markers may disrupt the natural flow of argumentation, imposing unnecessary cognitive load on readers.

The use of interactional metadiscourse presents a more complex picture. The use of hedges and boosters shows a marked asymmetrical distribution: the occurrence frequency of boosters (e.g., "clearly," "undoubtedly") is typically higher than that of hedges (e.g., "possibly," "perhaps"). This distribution preference may lead to a tendency towards absolutism in the tone of academic claims, weakening the capacity for appropriately reserving space for academic propositions, which contrasts with the customary cautious stance in academic writing.

In terms of authorial identity construction, the use of attitude markers is mostly confined to conventional expressions (e.g., "interestingly," "surprisingly"). The realization of nuanced practices that require reflecting the author's unique academic stance and evaluative perspective is relatively limited. The use of self-mentions also exhibits formulized characteristics, where the use of the first-person plural often replaces the singular form. While this choice aligns with the traditions of some disciplines, it may obscure the individual author's academic responsibility and innovative contribution. Overall, AI systems are still in an exploratory stage regarding balancing academic objectivity and the presentation of authorial voice. Their patterns of metadiscourse use reflect specific constraints of current technology in constructing complex authorial identities and handling the expression of uncertainty in academic discourse.

3. Discourse Style Negotiation and Fusion in Human-AI Interactive Writing Models

3.1 Iterative Adaptation Between Human Authorial Intent and AI Text Generation

In the context of human-AI collaborative academic writing, the formation of discourse style is fundamentally a continuous process of negotiation. This process begins with the human author's initial intent and evolves progressively through multiple rounds of prompts and feedback loops. The initial prompt typically contains basic content requirements and formal specifications. As the interaction deepens, the human author continuously refines their instruction strategy based on evaluations of the generated text. Such refinement may involve subtle adjustments to academic voice, precise control over argumentative density, or the explicit expression of rhetorical preferences, forming a distinctive phenomenon known as "prompt evolution"[4].

The quality of prompt engineering directly impacts the effectiveness of style negotiation. Simple instructions often trigger the AI system to invoke the most common academic expression patterns from its training data, leading to a generalized text style. In contrast, complex prompts rich in discipline-specific terminology, rhetorical structures, and expressive preferences can guide the system to generate texts with greater professional specificity and individual characteristics. During this process, the human author needs to develop a special "meta-prompt" capability, meaning the ability to diagnose the gap between the currently generated text and the ideal style, and to design targeted adjustment strategies.

This iterative adaptation mechanism creates a novel distributed system of writing cognition. The human author is responsible for providing disciplinary insight, stylistic guidance, and quality control, while the AI system undertakes the functions of language generation, pattern matching, and information integration. The continuous interaction between the two generates a unique creative dynamic: the author's initial intent may be refined, expanded, or even revised through dialogue with the AI system, while the AI's generative capacity is directionally stimulated under the author's guidance. The ultimately formed discourse style carries both the author's academic identity and incorporates the computational characteristics of the AI system, becoming a textual embodiment of human-AI intelligence fusion.

3.2 The Regulatory Effect of Disciplinary Paradigms on the Stylistic Output of AI-Generated Discourse

Disciplinary culture, as the deep structure of academic discourse, shapes the stylistic features of AI-generated texts through a dual pathway: the distribution of training data and interactive guidance. Texts in empirical scientific fields exhibit highly standardized structures, a preference for passive voice, and method-oriented exposition, aligning with the epistemological tradition of that domain which emphasizes objective verification. Interpretive disciplines, in contrast, demonstrate stronger argumentative features, theoretical speculation, and the use of the first person, reflecting their disciplinary traits that value stance articulation and meaning interpretation. This regulatory effect indicates that AI writing tools are, in essence, semiotic media embedded with specific epistemological assumptions.

The shaping influence of disciplinary conventions on discourse style is manifested not only at the macro-structural level but also permeates every dimension of micro-linguistic choice. Systematic differences exist across disciplines regarding citation norms, hedging strategies, and evaluative expression, and these differences are replicated to a considerable extent in AI-generated texts. For instance, in disciplines requiring highly cautious expression, AI texts correspondingly increase the

frequency of hedging devices; whereas in fields emphasizing certainty, the proportion of boosters rises noticeably.

When addressing research topics that are interdisciplinary or belong to emerging cross-disciplinary fields, AI systems demonstrate a degree of stylistic adaptability. They can adjust their discourse characteristics based on prompt requirements, attempting to integrate the discursive traditions of different disciplines. However, the coherence and naturalness of such stylistic blending largely depend on the breadth and quality of the relevant disciplinary corpora within the training data, as well as the precise guidance provided by human prompts. This regulatory role of disciplinary conventions demonstrates that AI-assisted writing is not a value-neutral instrumental entity, but rather a semiotic medium that carries the cultural genes and epistemological assumptions of specific disciplines. Its stylistic output is consistently constrained by the collective norms of an invisible disciplinary community[5].

3.3 The Tension Between Heterogeneity and Stylistic Consistency in Blended Texts

Academics texts produced through human-computer collaboration are, in essence, heterogeneous semiotic constructs that integrate the direct expression of human thought with computationally generated content from AI systems. This heterogeneity manifests across multiple levels of linguistic representation: at the lexical level, it appears as fluctuations in terminology density and distribution patterns; at the syntactic level, it presents as variations in structural complexity and sentence pattern diversity; and at the rhetorical level, it is displayed as differences in argumentation strategies and coherence mechanisms. These heterogeneous elements are not merely superimposed; rather, they form the distinctive stylistic features of blended texts through a complex integration process, creating a new form of academic discourse that differs from both purely human writing and entirely machine-generated text.

Maintaining stylistic consistency in blended texts presents unique challenges. The human author must keenly identify disjunctures between AI-generated segments and their own writing style. Such disjunctures may manifest as sudden shifts in argumentative rhythm, inconsistencies in metadiscourse patterns, or deviations in evaluative stance. Addressing these challenges requires authors to develop refined editing strategies, including the unification of terminological systems, the adjustment of sentence structures, the balancing of argumentative density, and the integration of metadiscourse modes. Successful stylistic fusion demands that authors possess a high degree of textual metacognitive ability, enabling them to accurately diagnose the root causes of stylistic inconsistency and implement targeted corrective measures[6].

Current AI systems exhibit structural limitations in sustaining personalized discourse styles. While a single prompt may generate text with local stylistic consistency, maintaining a stable reflection of deep-level academic identity characteristics across long-form text generation remains difficult. This tension of heterogeneity highlights the crucial role of the human author in text integration—they must simultaneously ensure content accuracy and stylistic consistency, acting as the ultimate stylistic arbitrator. The academic value of a blended text largely depends on the author's ability to effectively reconcile this tension, demonstrating rhetorical expressiveness while preserving internal coherence.

Conclusion

This study, through a systematic analysis of the impact of AI-assisted writing on the style of English academic discourse across the dimensions of lexis-grammar, rhetorical structure, and human-computer interaction, identifies distinctive stylistic features in AI-generated texts. At the lexical level, these features manifest as a concentration of core academic vocabulary and high-frequency use of nominalization structures. At the syntactic level, they are reflected in a preference for formulized structures and the passive voice. At the discourse level, they show a reliance on explicit cohesive devices while the deep coherence requires further enhancement. Regarding rhetorical structure, AI is capable of achieving standard move progression, yet its identification of research gaps and articulation of academic innovation remain insufficient. The process of stylistic negotiation within human-computer interaction further indicates that disciplinary paradigms and authorial intent jointly regulate the final stylistic orientation of the discourse. Meanwhile, the tension between heterogeneity and consistency in blended texts highlights the central role of the human author in stylistic mediation.

Future research could further explore the mechanisms for personalized style adaptation in AI

writing tools, deepen the comparative analysis of discourse styles across disciplines, and focus on the maintenance of stylistic consistency in long-form text generation. Concurrently, enhancing AI's support for implicit logic and creative expression within academic discourse through algorithmic optimization will be a significant challenge faced jointly by technological development and academic practice.

References

[1] Ma Yanhong. "Innovation in College English Teaching Models Empowered by Artificial Intelligence." Gansu Education Research .20(2025):13-18.

[2] Liu Fangyu. "Cultural Inheritance in High School English Grammar Teaching Empowered by Artificial Intelligence." Journal of the Chinese Society of Education S1(2025):15-17.

[3] Chen Cong, and Xie Zhilin. "The Application of AIGC Models Driven by Artificial Intelligence in the Practical Teaching of Basic Writing for English Majors." Journal of Hubei Open Vocational College 38.19(2025):158-160.

[4] Xiao Jingjing. "Construction of an Ethical Framework for AI-Assisted Writing Based on Qualitative Text Analysis." Chinese Journal of Scientific and Technical Periodicals 36.04(2025):449-459.

[5] Zhang Zihan. "Research on Opportunities, Challenges, and Coping Strategies for Students in AI-Assisted English Writing." Campus English .01(2025):46-48.

[6] Rong Rong. "Research on the Application of Natural Language Processing Technology Assisted Writing Based on Artificial Intelligence." Technological Innovation 11(2024):96-99.