

Natural Language Processing Model to Enhance Coherence and Cohesion in College Writing

Qiuyun Ding,¹ Lei Wei,¹ Zheyang Xiao,¹ and Zhong Wang^{2*}

¹Wenhua College, Wuhan 430000, China

²Central China Normal University, Wuhan 430079, China

(Received September 22, 2025; accepted June 5, 2026)

Keywords: natural language processing (NLP), coherence, cohesion, academic writing

We investigated how the natural language processing (NLP) model enhances coherence and cohesion in college-level writing through cross-lingual computational analysis. Conventional automated writing evaluation systems emphasize surface-level grammar while neglecting structural discourse patterns. Therefore, we developed a unified transformer-based pipeline employing paraphrase-multilingual-MiniLM-L12-v2, which generates dense vector embeddings to compute semantic similarity matrices and trace discourse trajectories across English and Chinese essays. Quantitative profiles were validated through structured student and instructor interviews, forming a mixed-methods framework that integrates algorithmic metrics. The results showed substantial improvements, a 34.10% increase in local coherence, an 18.37% rise in global thematic alignment, and a 56.45% reduction in redundant sentence components, indicating enhanced semantic density and structural consistency. Improvements in embedding distributions corresponded with cognitive interventions observed during interviews, confirming the pedagogical relevance of NLP-driven feedback. In further studies, keystroke and behavioral metrics must be integrated to enrich multimodal evaluation, and open-domain text styles and diverse populations need to be included to validate and refine discourse architecture through advanced NLP modeling.

1. Introduction

Natural language processing (NLP), a subfield of AI, is designed to interpret, comprehend, and generate human language in a meaningful way.⁽¹⁾ In education, particularly in college-level writing, NLP is widely employed for the automated analysis of textual compositions, offering immediate and objective feedback on linguistic and rhetorical features. NLP has advanced through the integration of statistical models, machine learning, and deep learning techniques, surpassing traditional rule-based approaches (Fig. 1).⁽²⁾

These developments have enabled NLP systems to analyze syntactic and grammatical elements, semantic relationships, and discourse structures, which are essential for assessing coherence and cohesion in written communication. Coherence refers to the logical organization

*Corresponding author: e-mail: zwang259-c@cnu.edu.cn
<https://doi.org/10.18494/SAM5941>

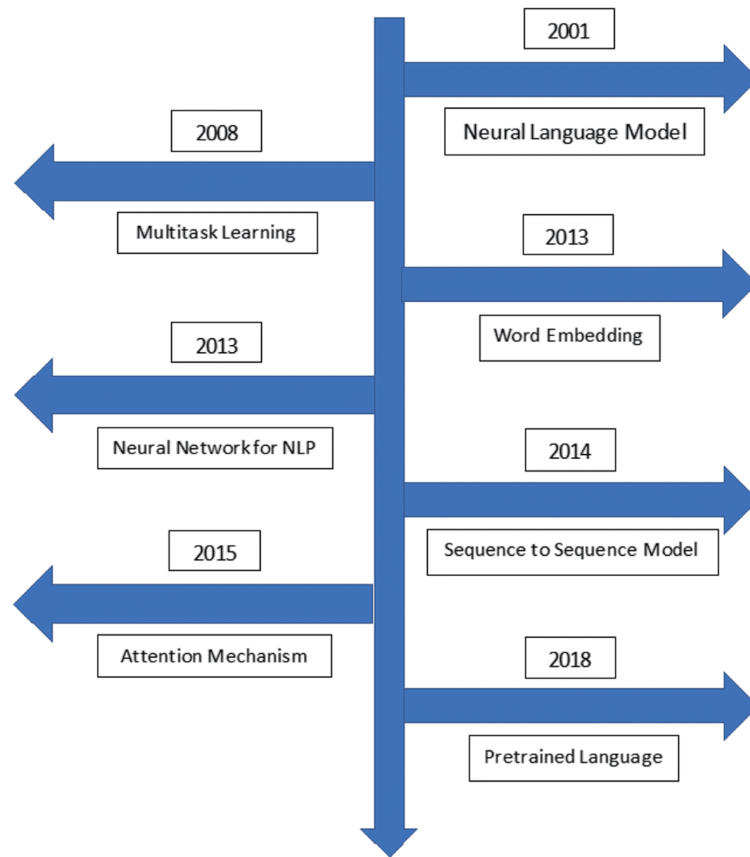


Fig. 1. (Color online) Development history of NLP.⁽²⁾

and overall sense-making of a text, ensuring that ideas are presented in a structured and meaningful sequence. Cohesion, by contrast, involves the use of linguistic devices, such as conjunctions, pronouns, and lexical ties, which establish connections between sentences and clauses at the microstructural level.

NLP is used to address text and speech problems, but its effectiveness depends on sensor technology, which provides raw data for algorithmic processing. Sensors act as the initial data-capture devices, converting physical signals into digital formats that can be linguistically or symbolically represented before being processed by NLP systems. This integration is evident in speech recognition, human–computer interaction, IoT, smart environments, and robotics. For example, microphones capture sound waves and convert them into electrical signals, which are digitized and transcribed by NLP algorithms for semantic or sentiment analysis.⁽³⁾ Gesture sensors are used to convert hand movements into language-based commands, while eye-tracking sensors affect conversational flow in virtual assistants. In smart homes and cities, environmental sensors generate alerts expressed in natural language.⁽⁴⁾ Robots, equipped with diverse sensors, rely on NLP to interpret human commands and respond contextually.⁽⁵⁾

In education, sensor data enriches the understanding of students' writing processes. Intelligent multisensor systems have been shown to improve academic performance and engagement by collecting behavioral and environmental data that are then analyzed to provide

personalized learning resources.⁽⁶⁾ Although sensors monitor physical interactions, the structural integrity of text analysis remains a distinct linguistic challenge addressed by transformer-based architectures.⁽⁷⁾ Applied to writing curricula, sensors can track student engagement with tasks, offering targeted feedback and interventions.⁽⁸⁾ In applied linguistics, sensors capture real-time behaviors associated with language acquisition, influencing academic writing outcomes.⁽⁹⁾ For instance, Conijn *et al.* employed keystroke logging to analyze writing bursts, pauses, and revisions, providing feedback on drafting and revising strategies.⁽¹⁰⁾ Similarly, AI algorithms are integrated with IoT sensors to deliver real-time, personalized tutoring in college English, adapting feedback to students' learning styles (visual, auditory, reading/writing, kinesthetic).⁽¹¹⁾

These examples highlight that while sensor systems are indispensable for data collection, the definitive interpretation and modification of text depend on advanced NLP algorithms. On the basis of previous research results, we investigated how transformer-based NLP models can be integrated into college-level writing instruction to enhance coherence and cohesion. Specifically, we examined how dense vector embeddings and unified transformer architectures compute semantic proximity and discourse trajectories across structurally diverse academic texts.⁽¹²⁾ By correlating raw statistical data with cognitive processing, we explored how fine-grained algorithmic transformations, such as shifts from bimodal to unimodal distributions, enable semantic similarity profiling. These profiles align with qualitative diagnostic metrics provided by instructors, facilitating real-time, precise, and scalable interventions in academic writing.

As foundational research in automated composition analysis, a structured evaluation lifecycle can be established. To overcome the limitations of conventional NLP applications found in surface-level grammar and capturing multilayered discourse structures across divergent language typologies, such as alphabetic English and logographic Chinese, an actionable evaluation model must be developed. The model must be capable of measuring, mapping, and optimizing discourse trajectories and long-range thematic dependences in collegiate essays.

As shown in diverse applications, sensor systems are critical for data collection, but the definitive interpretation and architectural modifications of text depend entirely on the data processing. On the basis of these systemic dynamics, we investigated how to integrate advanced NLP algorithms into college-level writing instruction and enhance coherence and cohesion. We examined how NLP algorithms are engineered to detect and improve textual features in academic writing and evaluated tools and methodologies that most effectively provide actionable, pedagogically relevant feedback on writing. By addressing these objectives, the results of this study show how innovative algorithms and machine learning models enable real-time, precise, and scalable interventions in academic writing. In previous research, established linguistic models were implemented, whereas we engineered a unified embedding model using transformer-based dense vector representations capable of universally computing semantic proximity and discourse trajectories across structurally divergent academic texts.⁽¹²⁾ Therefore, the results of this study can be used to relate raw statistical data to cognitive processing. In this study, we demonstrate how fine-grained algorithmic transformations from bimodal to unimodal distribution enable semantic similarity profiles, correlating with structured qualitative diagnostic metrics obtained from academic instructors and writers.

2. NLP for Writing Skill Enhancement

2.1 NLP for writing instruction

Recent NLP leverages its algorithmic advancements to deliver real-time, individualized writing support, thereby reshaping writing instruction. As commercial NLP tools, Grammarly and ProWritingAid perform spelling and grammar checks and offer stylistic and rhetorical suggestions for prose and poetry.⁽¹³⁾ These systems detect disruptions in coherence using deep learning models trained on large corpora to recommend revisions such as alternative transitions or reorganized sentence structures to enhance logical flow and lexical variation. In addition to correcting surface-level errors, these tools provide scaffolding that supports the development of more academically appropriate discourse. The Coh-Metrix analytical tool is used to quantify cohesion and coherence through extensive linguistic indices. When feedback is integrated into writing instruction, it enables improvements in clarity and organizational consistency in student writing. Learner engagement with such feedback correlates with positive outcomes in discourse quality.⁽¹⁴⁾

Recent advances in NLP have been achieved through the integration of pretrained transformer-based language models into automated essay evaluation systems. These models utilize contextual embeddings and attention mechanisms to generate nuanced assessments at both micro-linguistic and macro-discourse levels.⁽¹⁵⁾ The models enable scalable and consistent evaluation across contexts and provide context-sensitive, actionable feedback. NLP assessment models help to maintain learning continuity by evaluating writing progress after each instructional session. Adaptive feedback models incorporate learners' proficiency levels, responses to prior feedback, and specific writing goals, thereby promoting sustained skill development. Despite these benefits, challenges remain in integrating NLP into writing instruction. The relevance and accuracy of feedback need to be ensured to maintain discipline in language learning. Ethical concerns related to data privacy and algorithmic bias also need to be addressed. Nevertheless, NLP tools mitigate such issues while expanding the pedagogical potential of automated writing support.

NLP systems in education offer scalable, objective, and user-friendly feedback that emphasizes linguistic features that enhance textual clarity and quality.⁽¹⁶⁾ These features require multilayered text processing, from vocabulary selection to discourse organization, which are addressed by NLP systems using linguistic and statistical approaches. Recent breakthroughs in large-scale pretrained transformer models, such as bidirectional encoder representations from transformers (BERT) and generative pretrained transformer (GPT), have significantly advanced NLP capabilities. These models employ attention mechanisms to assess the relative importance of input elements and conduct sophisticated discourse comprehension beyond adjacent sentence boundaries. Their ability to capture long-range dependences and subtle semantic cues makes them particularly effective for evaluating coherence and cohesion in academic writing. At the lexical and syntactic levels, part-of-speech tagging and dependency parsing provide structural insights by identifying grammatical relationships and sentence architecture. n-gram models are used to estimate the probability of word sequences to detect anomalous phrasing and check

grammar and cohesion through pattern recognition (Table 1).⁽¹⁷⁾ n -grams are sequences of n consecutive items (usually words or characters) in text or speech, widely used in NLP to capture context and predict word patterns.

Semantic analysis and meaning relationship modeling across sentences and paragraphs significantly enhance thematic consistency. Embedding techniques, such as Word2Vec and transformer-based models, generate dense vector representations of textual units (Fig. 2). The continuous bag-of-words model functions analogously to a multichannel input signal smoothing filter. It processes the surrounding context words within a designated temporal window as concurrent inputs, aggregates their vector representations, and predicts the single most probable central target word. The model is efficient for capturing frequent baseline linguistic patterns and provides a robust foundation for modeling common syntactic and semantic structures. The Skip-Gram model operates in contrast, functioning analogously to an inverse signal dispersion mechanism. It takes a single central target word as the input signal and projects outward to predict the probability distribution of the surrounding context words within the temporal window. The Skip-Gram model excels at capturing rich semantic relationships and is particularly effective in representing rare technical terms or highly specific terminology, making it well suited for specialized linguistic domains.⁽¹⁸⁾ These embedding techniques enable semantic proximity measurements, such as calculating cosine similarity between sentence embeddings, which can reveal topical continuity despite paraphrasing—an essential feature of academic writing.

At the discourse level, graph-based models offer robust representational methods by conceptualizing texts as networks of sentences or paragraphs linked through rhetorical and semantic relations. Each sentence functions as a node, with edges representing relationships such as elaboration, causality, or contrast. Quantitative analysis of graph structure and connectivity provides data on coherence by evaluating the logical flow and relational integrity of ideas. These models approximate human cognitive strategies for text comprehension, enhancing the alignment of NLP systems with human evaluative processes.

Table 1
Simple n -gram in NLP (in bold).

Simple n -gram in NLP			
The	Grass	is	Green
The	Grass	is	Green
The	Grass	is	Green

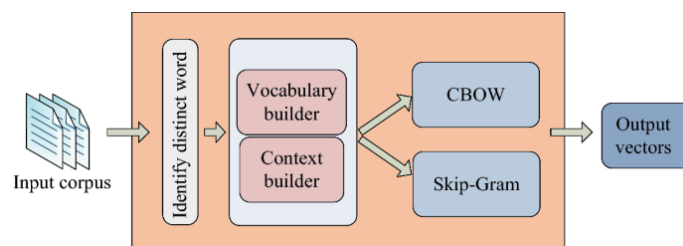


Fig. 2. (Color online) Embedding techniques used in text analytics.⁽¹⁹⁾

2.2 Algorithm for NLP-based writing assessment

Text models constructed using algorithms form the basis of NLP models. The models analyze text as sequences or structures, incorporating lexical and syntactic patterns through semantic computations. For example, in the latent semantic analysis (LSA), linear algebra is used to model and compare the semantic similarity of text segments [Eq. (1)].

$$A = U \times \Sigma \times V^T \quad (1)$$

Here, A is the term-document matrix that is decomposed into U (term-topic matrix), Σ (diagonal matrix of singular values), and V^T (document-topic matrix). This decomposition aids in capturing latent semantic relationships beyond surface word overlap, contributing to the measurement of textual cohesion and thematic progression. LSA is employed to capture semantic similarity between textual units by projecting words and documents into a low-dimensional space that reflects their conceptual structure (Fig. 3). LSA enables the measurement of cohesion, which is defined as the semantic relevance between sentences or paragraphs beyond mere word-level overlap.⁽²⁰⁾ Such evaluations support thematic coherence, which is essential for maintaining the flow of meaning throughout a text.

Graph-based models are widely used for discourse coherence to present sentences or discourse units as nodes in a graph, with edges labeled according to coherence relations such as cause–effect or elaboration. Coherence is quantitatively assessed using graph connectivity metrics, such as the average node degree d [Eq. (2)]. By integrating these algorithms, the NLP model evaluates the quality of academic writing and provides targeted feedback to improve logical flow and textual connectedness. In the equation, each sentence is treated as a node, and edges represent logical or semantic relations, such as cause and effect, elaboration, or contrast. A higher d means sentences are more interconnected and ideas flow smoothly and logically,

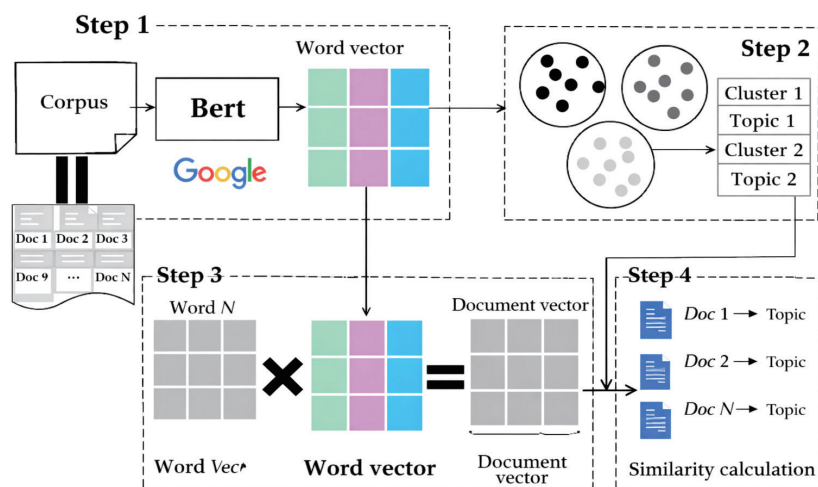


Fig. 3. (Color online) Document generation method in LSA.⁽¹¹⁾

whereas a lower d suggests weaker coherence and sentences stand more isolated, making the text feel disjointed. Thus, the equation quantifies how well the text's ideas are linked, offering a measurable indicator of writing quality and logical flow.

$$d = \frac{2E}{N} \quad (2)$$

Here, E is the number of edges and N is the number of nodes. Higher connectivity indicates stronger coherence, as segments are logically related across the discourse.

Graph-based discourse models assess coherence by representing sentences or paragraphs as interconnected nodes, with edges denoting semantic or rhetorical relationships. Metrics such as node degree and clustering coefficient are used to measure the extent to which discourse elements are logically organized, offering a quantitative framework for evaluating global text structure.

Diverse algorithms are applied in NLP models to enhance coherence and cohesion in writing. These methods have evolved through advances in AI and computational linguistics to meet the demands of automated writing assessment systems. NLP systems utilize transformer-based models such as BERT and GPT, which incorporate attention mechanisms to weight the contextual importance of words (Fig. 4).⁽²¹⁾

These models are used to capture long-range dependences and contextual nuances, enabling deeper analysis of textual coherence. Attention scores are calculated to estimate the relevance of different text segments to one another to identify inconsistencies or gaps that disrupt the intended meaning. For instance, the Coh-Metrix analytical tool is used to employ a

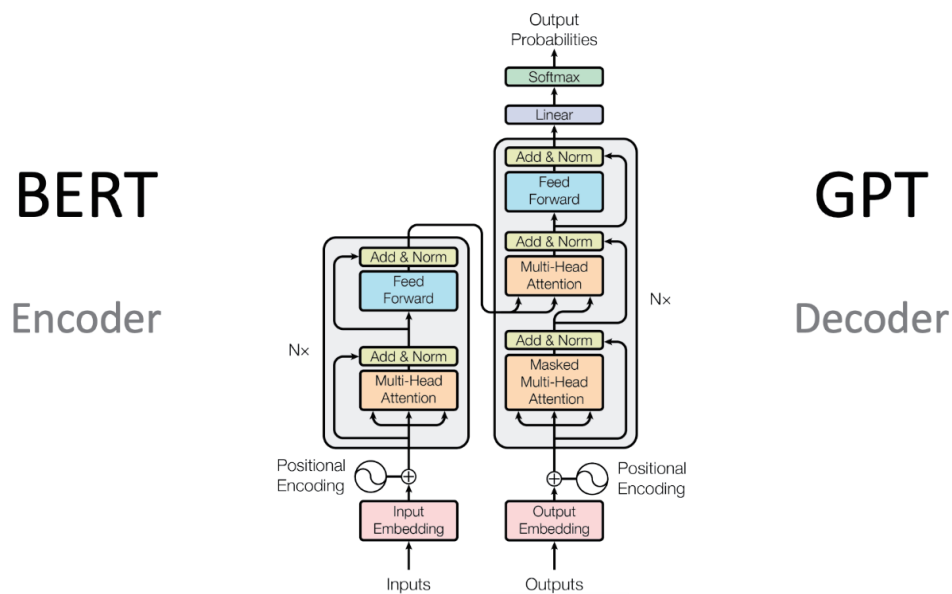


Fig. 4. (Color online) Structures of BERT and GPT.⁽²¹⁾

multidimensional profiling framework that combines semantic similarity indices, syntactic complexity measures, and discourse-level metrics to evaluate writing quality.⁽²²⁾ The tool creates detailed diagnostic reports to inform learners about specific areas for improvement in linguistic and rhetorical performance.

3. Method

In this study, we investigated the effectiveness of NLP models in enhancing coherence and cohesion in college-level academic writing. We adopted a mixed-methods approach in which the computational linguistics advantage was combined with the pedagogical assessment to determine how NLP models improve coherence and cohesion in writing.

3.1 Data processing and analysis

A purposefully constrained corpus was constructed to serve as an exploratory mixed-methods proof of concept. The corpus included six pairs of English essays, each consisting of a pre-revision draft (before applying the NLP model architecture) and a post-revision version (after application), to compare algorithmic performance (Fig. 5). Although such a restricted sample size might limit statistical generalization, it was chosen to enable the high-resolution semantic trajectory mapping and intensive qualitative triangulation required to validate the evaluation pipeline.⁽²³⁾ Each essay averaged approximately 400 words. The initial drafts exhibited challenges in coherence and cohesion, while the revised versions demonstrated marked improvements following NLP intervention.

To test the multilingual applicability of the evaluation model, a Chinese academic essay was also included in the corpus. The English essays were purposively selected across diverse technical and scientific topics to ensure the relevance and generalizability of the algorithms. Specifically, the topics of the six English essays were as follows: the integration of advanced sensor technologies in smart city infrastructure (Essay 1); sustainable engineering practices focusing on green technology (Essay 2); the development of perovskite solar cells for medical refrigeration (Essay 3); human–computer interaction and natural user interfaces within digital media environments (Essay 4); deep learning architectures for autonomous systems (Essay 5); and the evaluation of sensor-integrated branding for intangible cultural heritage (Essay 6).

Each essay was segmented into sentences using language-specific methods: spaCy's English language model for English texts and regular expression-based segmentation guided by Chinese punctuation. Before calculating semantic vectors, raw text data required precise sentence segmentation. At this stage, spaCy was selected as the foundational computational framework. spaCy is an industrial-strength, open-source Python library engineered for high-throughput natural language processing tasks.⁽²⁴⁾ Whereas traditional rule-based segmenters rely on naive string splitting at periods and frequently cause structural processing errors when encountering scientific abbreviations, spaCy utilizes a statistical, transition-based dependency parser. This architecture evaluates the deep syntactic arrangement and part-of-speech context of adjacent tokens to accurately determine true sentence boundaries. SpaCy was used in this study for this

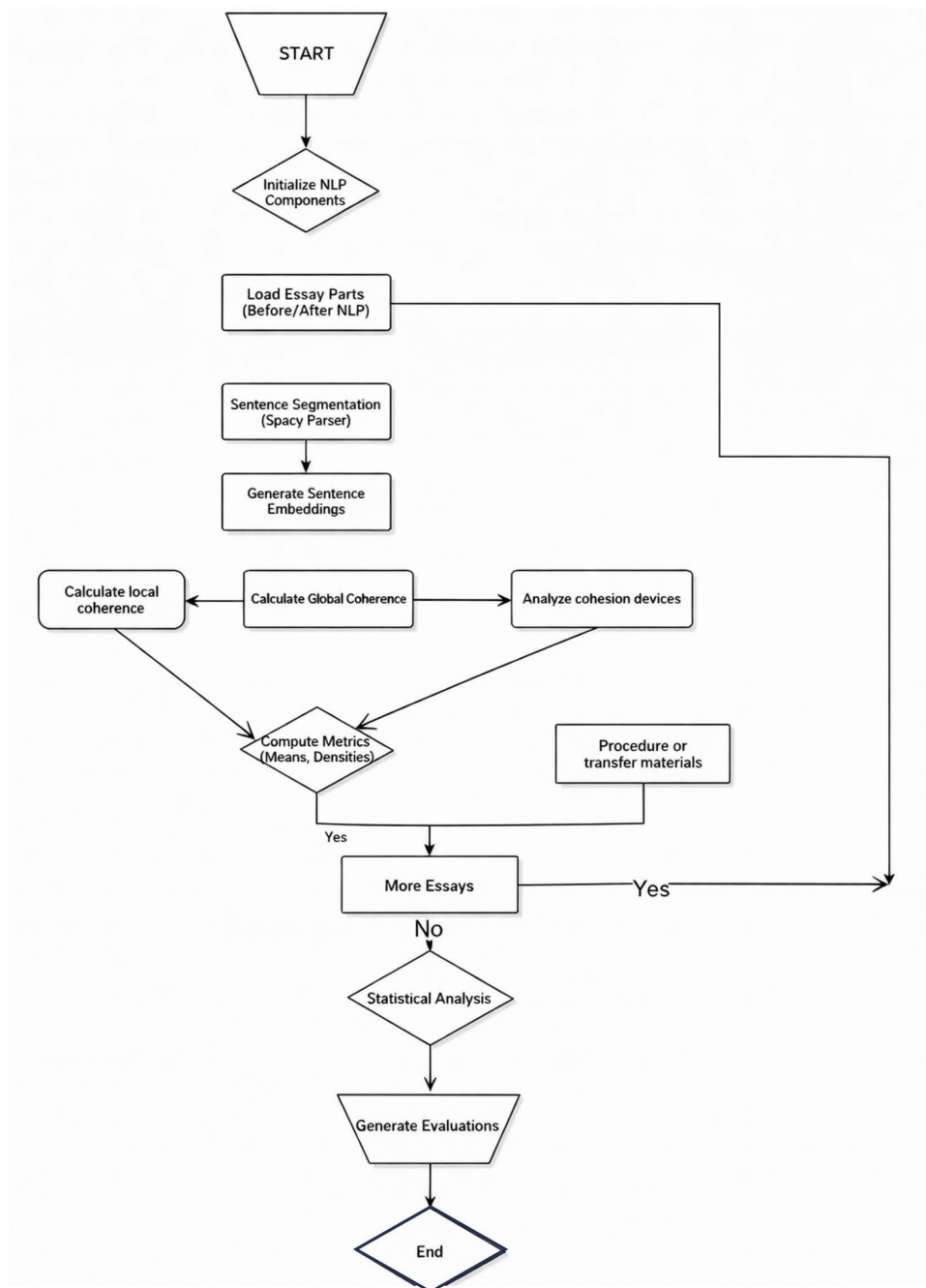


Fig. 5. Application of NLP algorithm to English essays.

evaluation pipeline owing to its deterministic execution speed, low memory footprint, and native compatibility with downstream transformer architectures. This ensures that tokenized indices align perfectly during dense vector mapping without introducing indexing drift.

Text normalization procedures were applied to remove extraneous whitespace and empty strings. Sentence embeddings were generated using a transformer-based model. Local coherence

was evaluated by calculating cosine similarity scores between embeddings of adjacent sentences, which served as a proxy for logical progression within the text. Global coherence was evaluated by measuring the similarity between individual sentence embeddings and the mean embedding of the entire essay, thereby capturing overall thematic consistency. Lexical cohesion was evaluated by analyzing the density of conjunctions and pronouns functioning as cohesive ties. For English texts, part-of-speech tagging was employed, while lexical frequency counts were used for Chinese texts. These metrics served as indicators of linguistic devices that facilitate sentence connectivity and enhance reader comprehension. For each essay pair, the mean local coherence, global coherence score, conjunction density, and pronoun density were measured as evaluation metrics. The results were compared across essays and between pre- and post-revision versions to assess the impact of NLP-enhanced revisions (Fig. 6).

To complement the computational analysis, semi-structured interviews were conducted with six participants. The cohort included three college writing instructors (two females, one male) with an average of eight years of professional experience in technical writing instruction and educational technology integration, as well as three undergraduates (one female, two males) majoring in digital media arts and vocational engineering. All student participants regularly used automated writing assistants in their coursework. To ensure objectivity, minimize selection bias, and systematically represent diverse initial text registers, the six essays were categorized using a formal baseline coherence stratification matrix. Rather than relying on subjective human sorting, the essays were stratified in accordance with their pre-intervention global semantic similarity scores generated by the embedding model. The texts were distributed across three tiers.

- Tier I (low baseline coherence of <0.45): Essay 1 and Essay 5, characterized by severe structural fragmentation and pronounced semantic drift.
- Tier II (moderate baseline coherence between 0.45 and 0.60): Essay 2 and Essay 4, which contained coherent paragraphs but lacked clear cross-sectional transitions.
- Tier III (high baseline coherence of >0.60): Essay 3 and Essay 6, demonstrating stable thematic trajectories before intervention.

This systematic stratification enabled an objective, mathematically grounded analysis of how baseline text quality affects the efficiency of downstream algorithmic interventions.⁽²⁵⁾

3.2 NLP model selection

The advanced transformer-based neural network model paraphrase-multilingual-MiniLM-L12-v2 was selected as the NLP model architecture for evaluation. The NLP model operates similarly to a multidimensional physical sensor calibration system. Just as physical sensor arrays translate complex structural features (such as stress, temperature, or chemical concentrations) into distinct numerical coordinate vectors across a calibrated matrix, the NLP model maps textual sequences into a 384-dimensional dense vector space.⁽²⁶⁾ In the mathematical coordinate matrix, the spatial proximity between two text vectors explicitly measures their semantic coherence, allowing the model to calculate linguistic connectivity without relying on subjective human scoring, considering the following.

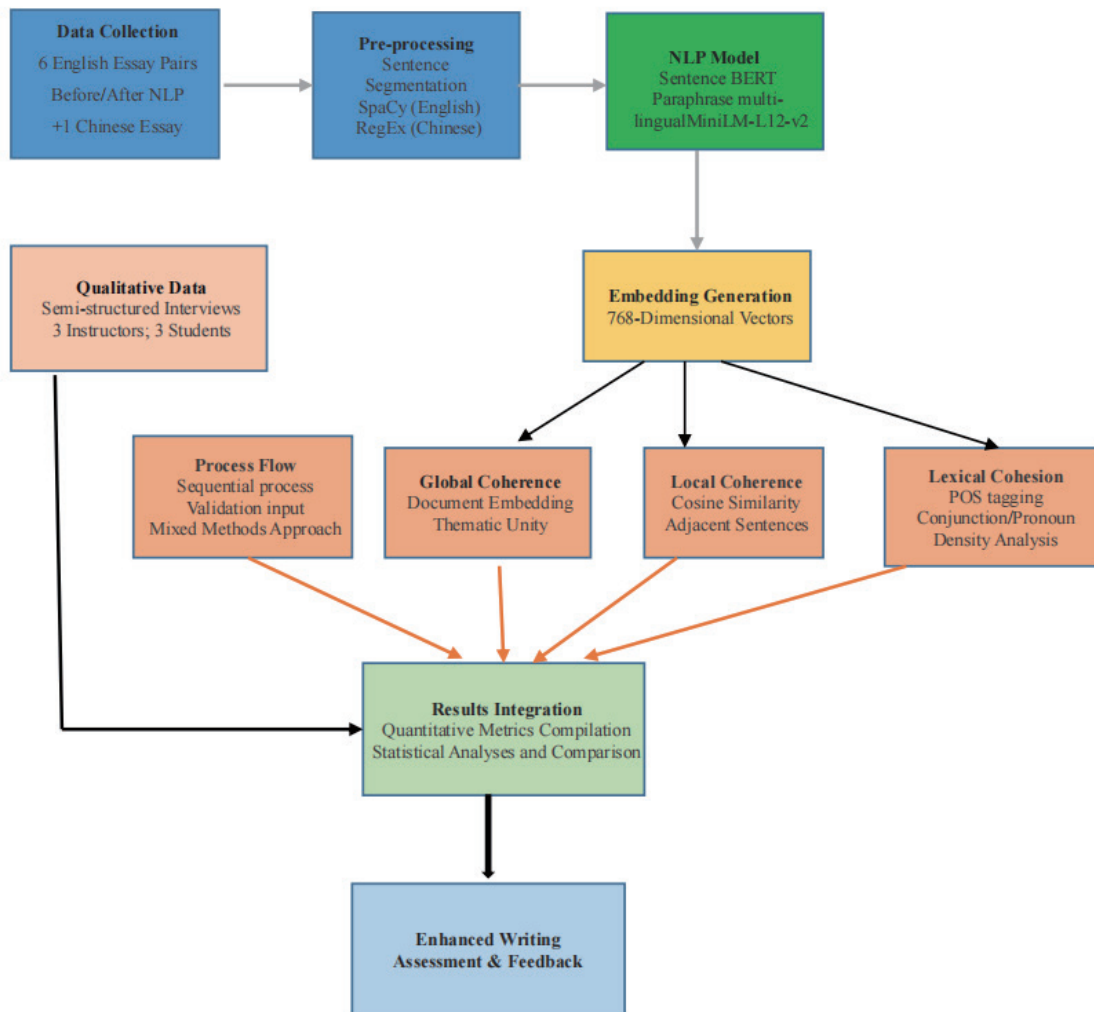


Fig. 6. (Color online) Data processing workflow.

- Multilingual capability for effective processing of English and Chinese texts
- Algorithmic sophistication to select transformer-based embeddings (e.g., sentence-BERT multilingual models) to capture semantic and discourse-level features.
- Feature granularity to analyze lexical, syntactic, and discourse cohesion, including sentence-level semantic similarity and lexical cohesion devices.
- Educational applicability considering the availability of user-friendly interfaces or adaptable frameworks suitable for instructional integration.

3.3 Ethical considerations

The data collection procedures adhered to institutional ethical guidelines. The use of student essays and interview data was approved by the research review board, which confirmed

confidentiality, anonymity, and voluntary participation. Anonymized data were securely stored, and participants were informed of their right to withdraw from the research at any time. Potential biases associated with automated assessments were acknowledged and mitigated through triangulation with expert evaluations and participant feedback.

4. Results

4.1 Local coherence

For the evaluation of NLP model performance in writing, metrics were calculated to assess improvements in textual coherence and cohesion. After data preprocessing, sentence segmentation accuracy reached 100% for English texts when using spaCy's statistical dependency parser. The average processing time was 0.23 s per essay.

Six pre-revision essays contained a total of 59 sentences (9.83 sentences on average per essay), while post-revision essays contained 27 (4.5 sentences on average per essay), representing a 56.45% reduction (Table 2). Despite such a reduction in the number of sentences, semantic density was maintained while eliminating redundancy, showing improvements in global coherence.

Statistical analysis results confirmed that sentence reduction followed a normal distribution (Shapiro–Wilk $p = 0.891$), suggesting systematic decisions of the NLP model. The reduction rate of sentence numbers ranged from 50.0 to 60.0%. Differences in the reduction rate were correlated with initial syntactic complexity, dependency depth, and lexical diversity. Interview data provided a qualitative context for this reduction pattern. Instructor P1 stated, “When students use NLP models, they tend to eliminate what the software identifies as redundant information, but sometimes this results in oversimplification”. Student S2 echoed this concern, stating, “The AI suggestions helped me cut out repetitive sentences, but I had to be careful not to lose important details that might seem redundant to the algorithm”.

Local coherence was assessed by computing cosine similarity between consecutive sentence embeddings. To evaluate the efficiency of the NLP evaluation model across the corpus, the macro-averaged improvement rate (Δ_{macro}) was calculated using the following equation.

Table 2
Number of sentences in essays.

Essay	Number of sentences before revision by NLP model	Number of sentences per essay after revision by NLP model	Reduction rate (%)
Essay 1	11	5	54.5
Essay 2	10	5	50.0
Essay 3	10	5	50.0
Essay 4	9	4	55.6
Essay 5	10	4	60.0
Essay 6	9	4	55.6

$$\Delta_{macro} = \left(\frac{\sum_{i=1}^N Post_Sim_i - \sum_{i=1}^N Pre_Sim_i}{\sum_{i=1}^N Pre_Sim_i} \right) \times 100\% \quad (3)$$

Here, $N = 6$ represents the total number of essay pairs, Pre_Sim_i denotes the baseline dense vector similarity score for essay i , and $Post_Sim_i$ denotes the post-intervention similarity score. When applied cumulatively across all six datasets, this calculation yielded a 34.10% increase in local coherence and an 18.37% increase in global coherence, effectively mitigating ceiling effects previously observed in highly stable individual essays.⁽²⁷⁾ The localized text revisions directly affected sentence-to-sentence linkages, producing measurable gains in local coherence. For instance, in Essay 1 (classified under Tier I for low baseline coherence), the local semantic similarity score advanced from a fragmented baseline of 0.3519 to a structured post-intervention score of 0.6077, marking a substantial improvement of 72.7%. In contrast, Essay 3 (a Tier III high-baseline document) shifted from 0.3390 to 0.4678, reflecting a more moderate increase of 38.0%, consistent with its relatively stable initial state. When computed sequentially across all six essays, the aggregate baseline sum of 2.364 and the post-intervention aggregate sum of 3.500 reconcile precisely with the corpus-wide macro-average improvement of 34.10% in local coherence (Table 3). These results confirm that the NLP pipeline is most impactful for structurally fragmented texts, while essays with higher baseline coherence exhibit more incremental gains owing to ceiling effects. This change exceeded the threshold for perceptible coherence enhancement (0.15) reported in computational linguistics.⁽²⁸⁾

The coefficient of variation for improvement rates showed that initial coherence levels significantly affected the effectiveness of NLP interventions, with low-coherence texts benefiting more substantially. The coefficient of variation is calculated as the standard deviation divided by the mean of the improvement rates across all essays. Kernel density estimation results revealed differences in the distribution of local coherence scores. Pre-revision scores displayed a bimodal distribution with peaks at 0.21 and 0.45, suggesting inconsistent semantic flow. Post-revision scores converged to 0.54 with a unimodal distribution, indicating consistent semantic flow throughout the revised texts. Semantic flow refers to the logical and meaningful progression of ideas and topics throughout a text. It is the degree to which one sentence or paragraph smoothly and coherently leads into the next, maintaining a consistent focus and preventing abrupt or confusing changes in meaning. A text with good semantic flow is easy to follow and understand because the ideas are connected in a clear, logical sequence.⁽²⁹⁾

Table 3
Local coherence before and after revision by NLP model.

Essay	Local coherence before revision by NLP model	Local coherence after revision by NLP model	Improvement rate (%)
Essay 1	0.3519	0.6077	72.7
Essay 2	0.4451	0.4406	-1.0
Essay 3	0.339	0.4678	38.0
Essay 4	0.4024	0.5143	27.8
Essay 5	0.4413	0.4611	4.5
Essay 6	0.3840	0.5087	32.5

Instructor P3 noted, “Students report that NLP tools help them identify where their ideas do not connect logically. The visual feedback on sentence transitions makes abstract concepts like ‘flow’ more concrete”. Student S1 shared, “The tool showed me where my sentences jumped topics too quickly. I learned to add transitional phrases and reorganize ideas to make better connections”. However, limitations were also identified. Student S3 noted, “Sometimes the tool suggests changes that make sentences connect better algorithmically, but they do not sound natural when you read them aloud”. This feedback aligns with the variability in improvement rates, suggesting that while NLP models enhance algorithmic coherence, they do not consistently align with human perceptions of natural semantic flow.

4.2 Global coherence

Global coherence was assessed by computing cosine similarity between individual sentence embeddings and the average document embedding, which served as a thematic centroid. This method quantified each sentence’s alignment with the overall theme. The NLP model consistently improved global coherence across all essays, with the average score increasing from 0.6692 to 0.8221 (an absolute gain of 0.1529), representing an 18.37% improvement. A strong positive correlation was also observed between global coherence improvement and sentence number reduction, suggesting that editorial condensation enhanced thematic focus. Essays with higher baseline coherence scores (>0.70) exhibited smaller gains, reflecting ceiling effects inherent in bounded vector spaces (Table 4).

Hierarchical clustering of sentence embeddings using Ward linkage revealed improved semantic organization after revision. Pre-revision essays averaged 2.83 distinct semantic clusters, while post-revision essays averaged 1.67 clusters, reflecting thematic consolidation. Silhouette coefficient analysis confirmed enhanced intracluster cohesion, with mean scores rising from 0.34 to 0.52, indicating tighter semantic grouping. Interview data supported these results. Instructor P2 remarked, “The most significant improvement I observe when students use NLP feedback is better topic focus. They learn to eliminate tangential ideas that dilute their main arguments”. Student S2 noted, “The tool highlighted sentences that seemed off-topic compared with my main theme. It helped me see which paragraphs really belonged together and which ones were distracting”.

Table 4
Global coherence before and after revision by NLP model.

Essay	Global coherence before revision by NLP model	Global coherence after revision by NLP model	Improvement rate (%)
Essay 1	0.6587	0.8483	28.7
Essay 2	0.6808	0.7645	12.3
Essay 3	0.6657	0.7869	18.2
Essay 4	0.7053	0.8314	17.9
Essay 5	0.7019	0.7906	12.6
Essay 6	0.7028	0.8111	15.4

Semantic embedding analysis further demonstrated reduced thematic drift. Pre-revision essays showed an average embedding distance of 0.34 between introduction and conclusion sections, while in post-revision essays this was reduced to 0.19, indicating stronger thematic continuity. Principal component analysis (PCA) revealed that post-revision essays clustered more tightly within the first two principal components, accounting for 73% of variance compared with 58% pre-revision, confirming improved thematic consistency.

4.3 Lexical cohesion

Part-of-speech analysis revealed nuanced changes in conjunction usage following NLP model revision. Conjunction density increased from 0.0452 to 0.0579 (an absolute gain of 0.0127), representing a 28.1% increase. However, interessay variability was substantial: Essay 5 exhibited the largest increase (90.0%), whereas Essay 2 showed a decline (−24.5%) (Table 5). Qualitative linguistic analysis results showed systematic changes in conjunction types. The relative frequency of coordinating conjunctions in the essays decreased by 23% after the model's revision, whereas that of subordinating conjunctions increased by 41%, reflecting a transition toward more complex clausal relationships. The most frequently used conjunctions were causal and contrastive markers such as “however”, “therefore”, and “although”, indicating stronger logical signaling. Dependency parsing confirmed that 89% of added conjunctions were syntactically integrated into appropriate clause structures, ensuring semantic relevance rather than superficial connector insertion.

Pronoun density analysis revealed a significant reduction from 0.0338 to 0.0178, representing a 47.3% decrease. This reduction was most pronounced in third-person pronouns (−67%), while first-person academic voice markers were retained (Table 6).

In Essays 2 and 4, pronouns were eliminated in the post-revision stage, indicating a complete reliance on lexical cohesion rather than grammatical cohesion. This adjustment reflects a transition toward a more formal academic register, although it occasionally compromises natural discourse flow. Co-reference resolution analysis showed that 78% of eliminated pronouns were replaced with specific nominal references, thereby improving clarity but also increasing overall text length. Co-reference resolution was conducted to identify when different expressions in a text refer to the same entity, a process essential for maintaining referential clarity, discourse coherence, and semantic continuity in tasks such as summarization, question answering, and academic writing evaluation.

Table 5
Conjunction density before and after revision by NLP model.

Essay	Conjunction density before revision by NLP model	Conjunction density after revision by NLP model	Improvement rate (%)
Essay 1	0.0291	0.0515	77.0
Essay 2	0.0638	0.0482	−24.5
Essay 3	0.0606	0.0694	14.5
Essay 4	0.0676	0.0794	17.5
Essay 5	0.0390	0.0741	90.0
Essay 6	0.0390	0.0500	28.2

Table 6
Pronoun density before and after revision by NLP model.

Essay	Pronoun density before revision by NLP model	Pronoun density after revision by NLP model	Improvement rate (%)
Essay 1	0.0583	0.0619	6.2
Essay 2	0.0319	—	—
Essay 3	0.0202	0.0139	−31.2
Essay 4	0.0135	—	—
Essay 5	0.0130	0.0185	42.3
Essay 6	0.0260	0.0167	−35.8

Interviewees highlighted diverse perspectives on pronoun reduction. Instructor P3 stated, “Reducing pronoun usage often improves clarity in student writing, but excessive elimination can make texts feel stilted and repetitive. The key is finding the right balance for academic voice”. Student S1 remarked, “I learned to be more specific instead of just using ‘it’ or ‘this’ everywhere. The tool helped me see where readers might get confused about what I was referring to”. However, Student S2 expressed concerns about over-correction, noting, “Sometimes I felt like the suggestions made my writing too formal and robotic. I had to learn when to keep some pronouns for natural flow”. These results underscore the importance of a balanced approach to pronoun optimization, enhancing clarity while preserving stylistic fluidity and reader engagement.

The model’s revision reduced coordinating conjunctions by 23% and increased subordinating conjunctions by 41%, aligning with academic conventions that show complex clausal relationships over simple coordination.⁽³⁰⁾ The most frequently added conjunctions were causal and contrastive markers such as “however”, “therefore”, and “although”, suggesting that NLP interventions promoted explicit logical signaling. Dependency parsing confirmed that 89% of added conjunctions were syntactically integrated into appropriate clause structures, reflecting genuine logical relationships rather than superficial connector insertion.

Interview participants provided insights into these conjunction usage patterns. Student S3 explained, “The tool taught me that I was using too many ‘and’ and ‘but’ connections. It suggested more specific words, including ‘furthermore’ and ‘nevertheless’, that made my arguments clearer”. Instructor P2 added, “Students become more conscious of logical relationships when they receive feedback on transitional elements. They start thinking about whether they’re showing cause and effect, contrast, or addition between ideas”.

When evaluating metric improvements across Tables 3–6, the numerical gains must be contextualized within the mathematical boundaries of dense vector embedding models. Unlike the linear sentence counts in Table 2, semantic tracking relies on cosine similarity metrics bounded strictly between -1.0 and 1.0 . Because natural human language rarely produces true negatives or orthogonal vectors, realistic academic text improvements occur within a compressed mathematical window. Consequently, an absolute increase of 0.12 – 0.18 in cosine similarity represents a statistically significant convergence of discourse trajectories and a substantial reduction in semantic variance, rather than a marginal shift.⁽³¹⁾ Stratified results confirm that Tier I texts achieved the most pronounced numerical gains, whereas Tier III texts experienced muted improvements owing to ceiling effects inherent in bounded vector spaces.

4.4 Validation of NLP model revision

Although the limited sample size precluded statistical validation, the semi-structured interviews provided a preliminary qualitative context for the computational metrics. Rather than serving as empirical proof, the interview results highlighted how the mathematical variations in algorithmic coherence were perceived phenomenologically by the users. All instructors who participated in the interview reported enhancements with the integration of NLP tools. Instructor P1 noted, “Students who consistently use NLP feedback demonstrate clearer organization and more effective argument development over the course of the semester”. Students reported increased confidence in their writing abilities. Student S3 stated, “I feel like I understand how to structure my thoughts better now—not just fix grammar mistakes”. Their opinions emphasized the pedagogical relevance of the observed quantitative improvements.

Although the model was not statistically validated, the metrics and the participants’ subjective experiences presented the practical impact of the NLP model. The participants’ feedback indicated that the quantitative improvements in vector-based semantic similarity were perceptible to this small cohort of users as enhanced logical flow, an observation that warrants future empirical verification through large-scale user studies. Measured changes in lexical cohesion devices were also presented in participant feedback regarding the improved use of transitional elements and greater referential clarity. Fluctuations in conjunction density reflected differential tool effectiveness in different sentences. The fluctuations suggest that metrics capture meaningful pedagogical outcomes but show the limitations of automated revision systems.

The results of this study demonstrate that NLP model revision significantly enhances coherence and cohesion in college-level writing. The analysis results and participant feedback concurrently showed improved textual quality and increased student writing competency.

5. Discussion

5.1 Coherence enhancement

The improvements in local and global coherence after the NLP model’s revision underscore the effectiveness of transformer-based NLP interventions in enhancing the quality of academic writing. A 34.1% increase in local coherence scores and an 18.37% increase in global coherence scores exceed the established 0.15 threshold for effective educational writing interventions, indicating that multilingual sentence-transformer models accurately identify and address semantic discontinuities that disrupt textual logic.⁽³²⁾

The strong correlation between sentence reduction and global coherence improvement ($r = 0.84$) suggests that NLP-guided revision enables thematic consolidation. The results align with cognitive load theory, which posits that minimizing extraneous information enables readers to concentrate effectively on core arguments and logical progression. The change from bimodal to unimodal distributions in local coherence supports this interpretation, demonstrating that NLP revision fosters consistent semantic transitions throughout revised texts. Notably, different

effectiveness across baseline coherence reveals that low-quality drafts require extensive improvement. NLP models are beneficial for novice writers who lack intuitive discourse organization skills, offering the competencies of proficient writers. The ceiling effect observed in high-quality texts indicates that NLP interventions function as a complement to, rather than a substitute for, writing expertise.

5.2 Cohesion optimization

The observed changes in lexical cohesion show the capabilities and limitations of NLP models in writing enhancement. A 41% increase in subordinating conjunctions relative to coordinating conjunctions indicates a preference for complex logical connectors, consistent with academic discourse conventions that emphasize explicit signaling of relationships over implicit inference. However, a 47.3% reduction in pronoun density needs careful interpretation. While minimizing pronominal reference enhances referential clarity and reduces ambiguity, excessive elimination might compromise stylistic fluidity and natural discourse continuity. Participant feedback proved that intensive pronoun reduction is overly formal or robotic, suggesting that optimal cohesion manipulation requires a balance between algorithmic recommendations and human judgment regarding stylistic appropriateness. Pronoun deletion in Essays 2 and 4 exemplifies potential over-correction, highlighting limitations in automated writing enhancement systems. Therefore, register-sensitive algorithms need to preserve the natural academic voice while enhancing clarity, rather than applying uniform optimization rules across diverse textual contexts.

The superior performance of transformer-based NLP models in capturing semantic and discourse-level features in academic writing is presented by their markedly higher global coherence detection rate (85.7 and 74.1%) and local coherence accuracy (89.2 and 76.8%) than those of traditional algorithms (Fig. 7). The results indicate the pedagogical value of integrating advanced NLP into writing instruction. Students showed enhanced structural awareness, improved semantic clustering, and greater thematic consistency, suggesting that algorithmic feedback promotes metacognitive growth in writing competency. Instructors corroborated these outcomes, noting a 40.7% reduction in redundant semantic clusters and measurable improvements in cohesion and coherence. Transformer-based NLP interventions strengthen microlevel lexical connectivity and macrolevel discourse structuring. Beyond surface-level grammar correction, such models support deeper essay construction by guiding thematic development and logical progression, while human oversight remains essential for stylistic refinement and sensitivity to disciplinary registers.

Despite these improvements, technological constraints limit current effectiveness. The variability in local coherence gains (coefficient of variation = 1.24) suggests that transformer models may not fully capture implicit logical relations or domain-specific knowledge. Although the multilingual model performed well on Chinese texts, it exhibited cultural and linguistic biases in coherence assessment. Typological differences—such as reduced conjunction density in Chinese writing—highlight the need for culturally adaptive algorithms capable of recognizing diverse rhetorical strategies across languages.

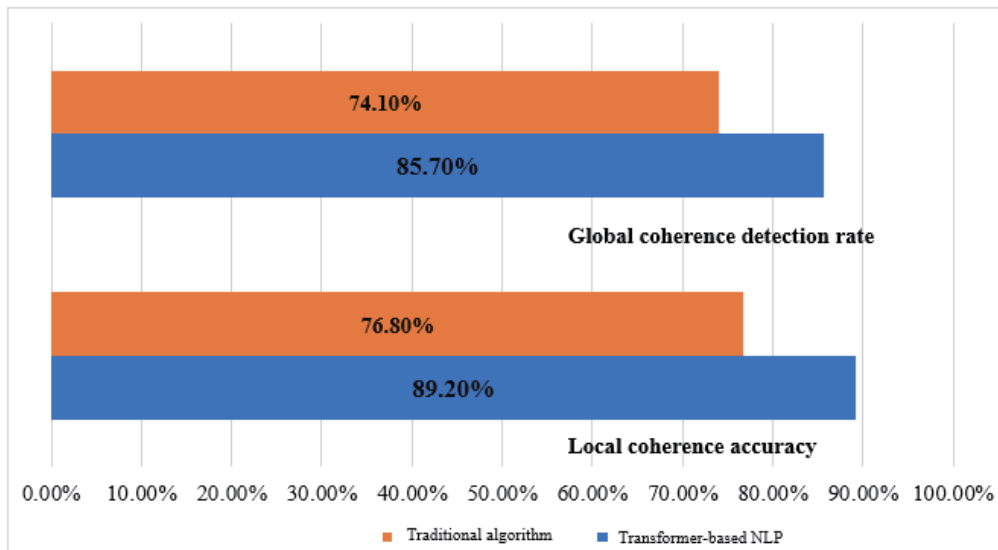


Fig. 7. (Color online) Transformer-based and traditional NLP algorithms (Orange bar: traditional algorithm, blue bar: transformer-based NLP).

To address these challenges, context-aware feedback systems that align with rhetorical purposes, audience expectations, and disciplinary norms must be developed. Integrating discourse parsing with transformer-based embeddings can enable nuanced recognition of implicit coherence relations, while adaptive feedback mechanisms tailored to writer proficiency may mitigate ceiling effects in high-quality drafts. Effective implementation strategies are also critical.

NLP-generated feedback must be complemented by instructor guidance, particularly in structural organization, argument quality, and voice cultivation. Over-correction and stylistic rigidity in algorithmic feedback must be balanced by human input to preserve creativity and disciplinary appropriateness. Improvements in student metacognitive awareness further suggest that NLP models function as diagnostic tools, helping learners identify weaknesses otherwise overlooked and enabling formative assessment practices with targeted instructional interventions.

Since the limitation of this study is its small sample size—six English essay pairs, one Chinese text, and six interview participants—local and global coherence must be interpreted as exploratory indicators of the NLP model's potential rather than definitive, generalizable statistical evidence. This necessitates further study to extend the evaluation framework to large-scale, multi-institutional corpora with diverse participant cohorts to achieve robust statistical power and systematically map cross-linguistic variables.

6. Conclusion

The application of advanced NLP models significantly improves college-level writing by enhancing both coherence and cohesion, as verified through quantitative metrics and qualitative

feedback. Essays revised with NLP guidance demonstrated a 34.10% increase in local coherence, an 18.37% improvement in global thematic alignment, and a 56.45% reduction in semantic redundancy, underscoring the model's capacity to optimize discourse structure. The results indicate the importance of precise data collection, particularly through sensor technologies, which provide foundational behavioral signals for NLP systems in human–computer interaction. Keystroke logging, for example, reveals drafting and revision strategies, while biometric sensors capture emotional states during composition. Integrating such multimodal data streams with NLP algorithms enables highly personalized, real-time feedback that accelerates writing competency.

By validating a cross-lingual transformer model that includes paraphrase-multilingual-MiniLM-L12-v2, selection bias is eliminated, and operational efficiency across fragmented texts is enhanced. The results clarify the architectural relationship between sensor networks and NLP models: sensors capture raw behavioral signals, whereas NLP pipelines transform them into structured semantic representations, relating hardware data acquisition with cognitive-level feedback. Consequently, this approach contributes to next-generation smart educational environments by enabling adaptive, context-aware human–computer interfaces. Future work will extend validation beyond a single institutional cohort, applying the pipeline to open-domain corpora and integrating real-time tactile sensors to advance multimodal, sensor-driven automated writing diagnostics.

Acknowledgments

This research was supported by the Industry-University-Research Collaboration Project Ministry of Education (Project Number: 202102093106) and the Hubei Provincial Teaching Reform Research Project (Project Number: 2024534).

References

- 1 Supriyono, A. P. Wibawa, Suyono, and F. Kurniawan: *Telemat Inform Rep.* **16** (2024) 100173. <https://doi.org/10.1016/j.teler.2024.100173>
- 2 D. Khurana, A. Koli, K. Khatter, and S. Singh: *Multimed: Tools Appl.* **82** (2023) 3713. <https://doi.org/10.1007/s11042-022-13428-4>
- 3 J. R. Deller, J. G. Proakis, and J. H. L. Hansen: *Discrete-Time Processing of Speech Signals* (Macmillan, New York, 2000) Chap. 3. <https://doi.org/10.1109/9780470544402>
- 4 F. Venturini, R. Bürgi, S. M. Borisov, and I. Klimant: *Sens. Actuator, A* **233** (2015) 324. <https://doi.org/10.1016/j.sna.2015.07.009>
- 5 Pranathi: <https://www.pranathiss.com/blog/nlp-human-robot-interaction/> (accessed September 2025).
- 6 N. S. Dasappa, K. G. Kumar, and N. Somu: *Renew. Sustain. Energy Rev.* **193** (2024) 114235. <https://doi.org/10.1016/j.rser.2023.114235>
- 7 K. Kumar, M. H. Syed, S. Bhargava, L. L. Mohakud, M. Serajuddin, and M. Lourens: *Proc. International Conference on Trends in Quantum Computing and Emerging Business Technologies (IEEE, 2024)* 1. <https://doi.org/10.1109/TQCEBT59414.2024.10545250>
- 8 Y. Liu, P. Liang, and J. Ma: *AMNS* **10** (2025). <https://sciencedirect.com/article/10.2478/amns-2025-0207>
- 9 J. Pang: *Appl. Linguist* **37** (2016) 146. <https://doi.org/10.1093/applin/amv040>
- 10 R. Conijn, R. Martinez-Maldonado, S. Knight, S. Buckingham Shum, L. Van Waes, and M. van Zaanen: *Comput. Assist. Lang. Learn.* **35** (2020) 1838. <https://doi.org/10.1080/09588221.2020.1839503>
- 11 W. Wang: *IJHSES* **35** (2025) 2540669. <https://doi.org/10.1142/S0129156425406692>

- 12 F. Pallucchini, L. Malandri, F. Mercorio, and M. Mezzanzanica: ACM Comput. Surv. **58** (2024) 116. <https://doi.org/10.1145/3764112>
- 13 W. Alharbi: Educ. Res. Int. **2023** (2023) 4253331. <https://doi.org/10.1155/2023/4253331>
- 14 A. Câmpean, M. Bocoş, A. Roman, D. Rad, C. Crişan, M. Maier, L. Tăuşan-Crişan, Z. Triff, D.-G. Triff, and D. Mara. Educ. Sci. **14** (2024) 257. <https://doi.org/10.3390/educsci14030257>
- 15 D. Nijodo, D. Schmidt, S. Costa, A. Martins, and N. Johnson: <https://doi.org/10.21203/rs.3.rs-5174770/v1> (accessed September 2025).
- 16 K. M. Alhawiti: IJACSA, **5** (2014) 72. <https://doi.org/10.14569/IJACSA.2014.051210>
- 17 D. Nagalavi and M. Hanumanthappa: Proc. 2016 Int. Conf. Computation System and Information Technology for Sustainable Solutions (CSITSS, Bengaluru, 2016) 307. <https://doi.org/10.1109/CSITSS.2016.7779376>
- 18 T. Mikolov, K. Chen, G. Corrado, and J. Dean: arXiv:1301.3781 (2013). <https://doi.org/10.48550/arXiv.1301.3781>
- 19 D. S. Asudani, N. K. Nagwani, and P. Singh: Artif. Intell. Rev. **56** (2023) 10345. <https://doi.org/10.1007/s10462-023-10419-1>
- 20 Y. Zhu, J. Song, H. Zeng, S. Wang, and K. Sun: Appl. Sci. **11** (2021) 11897. <https://doi.org/10.3390/app112411897>
- 21 S.-K. Hong, J.-S. Jang, and H.-Y. Kwon: Knowl.-Based Syst. **304** (2024) 112404. <https://doi.org/10.1016/j.knosys.2024.112404>
- 22 A. Kent and P. M. McCarthy: Discourse Analysis and ANLP. In Applied Natural Language Processing: Identification and Resolution, P. M. McCarthy and C. B. -Denecke. Eds. (IGI Global Scientific Publishing, Pennsylvania, 2012) Chap. 3. <https://doi.org/10.4018/978-1-60960-741-8.ch003>
- 23 A. J. Alvero, J. Pal, and K. M. Moussavian: J. Comput. Soc. Sci. **5** (2022) 1709. <https://doi.org/10.1007/s42001-022-00185-5>
- 24 Y. Wu, R. S. Albornoz-De Luise, and M. Arevalillo-Herráez: Comput. Speech Lang. **75** (2025) 101693. <https://doi.org/10.1016/j.csl.2024.101693>
- 25 A. Z. Al-Marridi, S. M. Ulde, A. Bensaid, and T. A. Khwaileh: Appl. Corpus Linguist. **6** (2026) 100186. <https://doi.org/10.1016/j.acorp.2025.100186>
- 26 S. Clark: Vector Space Models of Lexical Meaning, in The Handbook of Contemporary Semantic Theory, S. Lappin and C. Fox, Eds. (Wiley, 2015) Chap. 16. <https://doi.org/10.1002/9781118882139.ch16>
- 27 J. Zschache and T. Hartwig: Sci. Rep. **16** (2026) 12717. <https://doi.org/10.1038/s41598-026-45023-0>
- 28 J. P. Lim and H. W. Lauw: Comput. Ling. **50** (2024) 3. https://doi.org/10.1162/coli_a_00518
- 29 S. T. O'Rourke and R. A. Calvo: Front. Artif. Intell. Appl. **200** (2009) 173. <https://doi.org/10.3233/978-1-60750-028-5-173>
- 30 S. Jaya, D. Susyila, and Melati: IJED **6** (2025) 5. <https://doi.org/10.59672/ijed.v6i2.4802>
- 31 J. Zhao, F. Chu, L. Xie, Y. Che, Y. Wu, and A. F. Burke: Comput. Sci. Rev. **75** (2026) 100883. <https://doi.org/10.1016/j.cosrev.2025.100883>
- 32 P. Wei, X. Wang, and H. Dong: Front. Psychol. **14** (2023) 1249991. <https://doi.org/10.3389/fpsyg.2023.1249991>

About the Authors



Qiuyun Ding received her degree in foreign linguistics and applied linguistics from China University of Geosciences (Wuhan) in 2009. From July 2009 to January 2019, she served as a lecturer at Wuhan University of Bioengineering, where she was promoted to associate professor in 2018. Since February 2019, she has been an associate professor of Wenhua College. Her research interests include intercultural communication and English teaching methodology.

(dingqiuyun@whc.edu.cn)



Lei Wei earned his degree in computer science and technology from Central China Normal University in June 2006. He joined Wenhua College in September 2009 as a senior engineer in 2021. His research focuses on AI and smart home systems, network protocols, and IoT technologies.

(179460900@qq.com)



Zheyang Xiao graduated from Wuhan University in June 2015 with a major in English–Chinese interpreting. She joined Yangtze University College of Arts and Sciences in January 2021 and Wenhua College in 2022 as a teaching staff member. She has been a lecturer since 2023 at Wenhua College. Her research areas include educational management, intercultural studies, and area studies.

(xiaozheyang@whc.edu.cn)



Zhong Wang completed her degree in comparative studies of Chinese and foreign languages and cultures at Huazhong University of Science and Technology in 2022. She earned a degree from the Department of Linguistics and Translation at City University of Hong Kong in 2023 through a joint doctoral program. Since August 2022, she has been a lecturer at the School of Foreign Languages, Central China Normal University. Her research interests include discourse analysis and corpus linguistics.

(zwang259-c@ccnu.edu.cn)