ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

USING NATURAL LANGUAGE PROCESSING TO ADVANCE SECOND LANGUAGE WRITING ASSESSMENT: EVIDENCE FROM CORPUS-BASED RESEARCH

Tassawar Ali Khan
Assistant Professor, Government Degree College Kohsar, Latifabad Hyderabad
Email: ttasawar110@gmail.com
Mahnoor Shaikh
Lecturer (English), Govt. Nazareth Girls Degree College, College Education Department, Government of Sindh
Email: mahnoorshaikh187@gmail.com
Parwez Ali Bughio
Lecturer,GC University Hyderabad
Email: parwezali.bughio@gcuh.edu.pk

Abstract
This paper explores how Natural Language Processing (NLP) can support the assessment of second language (L2) writing, particularly in relation to textual coherence and cohesion. The study centers on the DECOR (Detect, Explain, and Rewrite) framework as a tool for identifying discourse-level weaknesses and generating revisions that improve the organization and connectedness of learner writing. Data for the analysis are drawn from the EF-Cambridge Open Language Database (EFCAMDAT), a large-scale learner corpus containing more than one million English texts produced across different CEFR proficiency levels. Because the corpus includes extensive learner metadata and draft histories, it offers a valuable basis for examining patterns of writing development over time. Methodologically, the research adopts a mixed approach that combines automated analysis and revision through DECOR, corpus-informed feature extraction, and human evaluation of writing quality. The effectiveness of NLP-based intervention is measured through comparisons between original and revised drafts, with particular attention to gains in coherence and lexical cohesion. The findings reveal common discourse-related difficulties among L2 writers and show that NLP-driven feedback can make a meaningful contribution to writing improvement. By connecting computational analysis with established human rating practices, the study underscores the educational value of AI-assisted writing assessment and promotes scalable, evidence-based, and learner-oriented methods for multilingual writing instruction.

Keywords: *Natural Language Processing (NLP), Second Language Writing, DECOR Framework, EFCAMDAT Corpus, Automated Writing Assessmen*

INTRODUCTION

Background

Assessment of second language (L2) writing has advanced considerably over the past two decades, yet important concerns remain regarding reliability, scalability, and instructional usefulness. Traditionally, writing evaluation has relied on human raters using holistic or analytic rubrics to judge features such as grammatical control, lexical sophistication, fluency, and discourse organization. Among these dimensions, coherence—understood as the meaningful and logical progression of ideas—has consistently been recognized as a central indicator of writing quality (Zhang et al., 2024). However, despite its importance, coherence remains one of the most difficult aspects to evaluate in a consistent and objective manner. Even trained instructors may

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

differ in their judgments because coherence is abstract and depends heavily on interpretation of discourse-level relationships.

With the growth of computational linguistics and natural language processing (NLP), researchers have increasingly explored the possibility of automating writing assessment. When such technologies are combined with large, annotated learner corpora, they create new possibilities for scalable and evidence-based analysis of learner performance. Earlier NLP-based assessment tools concentrated mainly on surface-level elements, including grammar, spelling, and vocabulary use (Mayormente & Gumpal, 2025). More recent innovations, particularly those driven by deep learning and advanced language models, have expanded this focus to include higher-level textual features such as cohesion, coherence, and argumentative organization (Koopman & Guardiano, 2022).

One promising development in this area is the DECOR framework, which stands for Detect, Explain, and Rewrite. This model is intended to assess coherence in L2 writing through a three-stage process: locating incoherent portions of text, explaining the causes of incoherence, and revising those segments to improve clarity and flow. Because of this layered design, DECOR offers more than simple error identification; it also provides constructive and potentially pedagogically useful feedback. Although the framework has shown encouraging results in smaller-scale studies and revision-based contexts, it has not yet been extensively examined using large learner corpora such as the EF-Cambridge Open Language Database (EFCAMDAT).

EFCAMDAT is one of the most extensive open-access learner corpora available, containing over one million writing samples produced by learners across all CEFR proficiency levels. In addition to the texts themselves, the corpus includes detailed metadata related to learner background, proficiency level, task type, and revision history. These features make it especially valuable for investigating patterns of L2 writing development and for testing the performance of NLP-based assessment systems on a broader scale. Despite its richness, relatively few studies have used EFCAMDAT to investigate discourse-level phenomena, particularly coherence, which leaves an important gap in both applied linguistics and NLP-oriented writing research.

Bringing together the DECOR framework and the EFCAMDAT corpus offers a strong foundation for improving writing assessment practices. This integration makes it possible to move beyond surface correction and toward deeper, pedagogically meaningful feedback on discourse organization. It also reflects recent directions in applied linguistics that encourage the use of AI-driven tools not only for scoring purposes but also as part of a formative learning process that promotes learner autonomy and long-term development (Zhang et al.,2024; Granger, 2021).

Research Objectives

• To examine how effectively the DECOR framework can identify and revise incoherent sections in learner texts drawn from the EFCAMDAT corpus.

• To determine the extent to which DECOR-based improvements in coherence correspond with human judgments across different CEFR proficiency levels.

Research Questions

1. How effectively can the DECOR framework identify and revise incoherent passages in learner writing at different CEFR levels?
2. To what extent do the coherence improvements produced by DECOR correspond to human evaluations of writing quality and coherence?

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Problem Statement

Although automated writing assessment has made notable progress, discourse-level evaluation—especially the assessment of coherence—remains underdeveloped in many existing systems. Most NLP-based tools continue to prioritize grammatical accuracy and lexical variation, offering only limited attention to higher-level discourse issues that strongly shape readers' understanding of a text. Human raters, by contrast, can recognize coherence-related problems, but they often do so inconsistently and cannot apply such judgments efficiently on a large scale. The absence of a reliable and scalable method for assessing coherence creates challenges for both research and pedagogy. By applying the DECOR framework to the large-scale EFCAMDAT corpus, the present study seeks to address this gap through a corpus-informed and systematic approach to evaluating and improving coherence in L2 writing.

REVIEW LITERATURE

The use of Natural Language Processing (NLP) in second language (L2) writing assessment has gained remarkable attention over the past decade due to rapid technological progress and the increasing demand for scalable, objective, and data-informed assessment methods. Traditional writing assessment has long depended on human raters, whose judgments, although valuable, often face limitations related to consistency, fairness, and efficiency. As a result, researchers have increasingly turned to computational approaches to support and improve writing evaluation processes. This literature review examines three closely connected areas that form the foundation of this study: Automated Writing Evaluation (AWE) systems, coherence in L2 writing assessment, and the role of learner corpora in corpus-informed assessment. Together, these areas provide both the theoretical and practical basis for applying coherence-focused NLP tools such as the DECOR framework in L2 writing assessment.

2.1 Automated Writing Evaluation (AWE) Systems

Automated Writing Evaluation (AWE) systems have undergone substantial development, moving from simple grammar-checking tools to more advanced platforms capable of offering both formative and summative feedback on writing quality. Earlier systems such as PEG and e-rater mainly focused on surface-level linguistic features including grammar, mechanics, vocabulary, and syntactic accuracy (Granger, 2021). These systems were effective for identifying local language errors, but they provided limited support for evaluating higher-level aspects of writing such as idea development and discourse organization.

As NLP techniques advanced, especially with the rise of machine learning and transformer-based models, AWE systems began to address more complex textual dimensions. Modern tools are increasingly able to evaluate coherence, cohesion, and rhetorical structure in addition to sentence-level correctness (Hanaoka & Izumi, 2021). This shift reflects a broader movement in language assessment away from discrete grammar testing and toward more communicative and holistic evaluation of writing ability.

Contemporary platforms such as Grammarly, Criterion, and WriteToLearn now offer feedback related to sentence variety, idea progression, and organizational structure, alongside traditional grammar correction. Some peer-review platforms also provide similar forms of support through collaborative feedback (Ariely et al., 2020). Despite these improvements, many commercial AWE systems still prioritize local error correction such as spelling and grammar over broader discourse-level concerns. As a result, coherence-focused feedback remains relatively underdeveloped.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Recent studies have attempted to overcome this limitation by integrating neural models trained on large annotated corpora. For example, GPT-based systems have enabled more detailed analysis of learner writing, particularly in tasks involving summarization, paraphrasing, and text restructuring (Abraha & Nazir, 2024). However, while these models offer powerful analytical capabilities, concerns remain regarding transparency, interpretability, and alignment with pedagogical goals in classroom contexts.

2.2 Coherence in L2 Writing Assessment

Coherence is widely recognized as one of the most essential characteristics of effective writing, yet it remains one of the most difficult features to define and assess systematically. In L2 writing, coherence problems often emerge because of limited vocabulary resources, weak control of discourse structures, and interference from the writer's first language (Hartwell & Aull, 2021). Human raters may be able to recognize when a text feels unclear or disconnected, but identifying the exact source of the problem and offering practical revision strategies requires a more structured approach.

To address this challenge, several NLP-based tools have been developed to measure coherence using indicators such as lexical overlap, semantic similarity, and entity transition patterns. Tools like Coh-Metrix and TAACO analyze cohesion by examining features such as referential overlap, causal connectives, and syntactic variation, which serve as indirect indicators of coherence (Delu, 2021). Although these tools are valuable for research purposes, they often generate numerical scores rather than detailed instructional feedback, limiting their usefulness in classroom writing development.

The DECOR framework represents an important advancement in this area because it combines detection, explanation, and revision within a single model. Zhang et al. (2024) introduced DECOR as a benchmark task for identifying incoherence in L2 English writing. The framework requires models to locate problematic discourse segments, explain why these segments are ineffective, and propose suitable revisions. Unlike earlier coherence tools, DECOR emphasizes pedagogically meaningful feedback and supports formative assessment practices commonly valued in applied linguistics.

Studies applying DECOR to learner writing have demonstrated its potential to capture discourse-level issues that are often missed by conventional assessment tools. Problems such as topic drift, semantic inconsistency, and abrupt transitions can be detected and revised in ways that closely reflect expert human judgments (Zhang et al., 2024). This makes the framework particularly useful in classrooms where teachers face time constraints, large student populations, or limited training in discourse-level feedback. In this sense, NLP functions as a supportive resource that strengthens human assessment rather than replacing it.

2.3 Learner Corpora and Corpus-Informed Assessment

Learner corpora have become an essential resource in second language acquisition (SLA) research and language assessment because they provide authentic samples of learner production that can be analyzed computationally. These corpora typically include texts written by L2 learners along with information about proficiency levels, error annotations, and revision histories, making them highly valuable for training and evaluating NLP models.

Among the most significant learner corpora is the EF-Cambridge Open Language Database (EFCAMDAT), which contains more than one million learner texts produced across a wide range of CEFR proficiency levels (Granger,2021). The size and detail of this corpus make it

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

particularly useful for examining patterns of language development across different stages of proficiency.

Researchers have used EFCAMDAT to investigate features such as lexical diversity, syntactic complexity, and error frequency, often connecting these features to CEFR levels and human rating outcomes (Koopman & Guardiano, 2022). More recent studies have employed neural embedding models to identify semantic and structural patterns in learner writing and improve the predictive accuracy of automated scoring systems (Hanaoka & Izumi, 2021).

Despite its richness, EFCAMDAT has been used less frequently for coherence-focused research. Most existing studies emphasize sentence-level accuracy or vocabulary growth, while discourse-level development remains comparatively underexplored. This creates an important research gap, particularly in understanding how learners organize ideas, manage transitions, and maintain thematic consistency across proficiency levels. Applying coherence-oriented models such as DECOR to EFCAMDAT may offer valuable insights into the development of higher-order writing skills.

In addition, corpus-informed feedback systems have shown strong potential for improving writing instruction. By identifying common learner difficulties and providing targeted support based on corpus evidence, these systems promote data-driven teaching and personalized learning pathways (Schmidt, 2022). Expanding such systems to include coherence assessment would significantly enhance their pedagogical value and better align automated feedback with higher-level writing instruction.

Research involving DECOR has further shown that discourse-level issues such as topic drift, semantic conflict, and weak transitions can be addressed more effectively when coherence analysis is integrated into writing support systems. These findings suggest that NLP can help teachers provide deeper and more meaningful writing feedback, especially in educational contexts where time and instructional resources are limited (Zhang et al., 2024).

2.4 Pedagogical Considerations and Future Directions

The educational value of NLP tools extends beyond diagnostic assessment to their role in supporting formative learning processes. Effective writing instruction depends heavily on revision, reflection, and meaningful engagement with feedback. Intelligent systems that explain problems and suggest revisions can strengthen these processes more effectively than tools that only highlight surface-level mistakes (Ariely et al., 2020).

However, several challenges must still be addressed. First, the reliability of coherence detection must be tested across diverse learner populations and writing contexts. Expectations of coherent writing may vary depending on genre, task type, age group, and cultural background, which means that adaptable and context-sensitive models are necessary (Delu, 2021).

Second, transparency remains a major concern. Learners and teachers need to understand how automated systems generate feedback and why certain suggestions are made, particularly in high-stakes assessment settings where fairness and trust are essential.

Ethical issues such as algorithmic bias, excessive dependence on technology, and data privacy must also be carefully considered. As NLP becomes more integrated into educational environments, ongoing evaluation and stakeholder involvement are necessary to ensure that these tools serve pedagogical goals rather than purely administrative functions.

Overall, existing literature strongly supports the integration of NLP-based tools such as DECOR into L2 writing assessment, particularly when combined with large learner corpora like

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

EFCAMDAT. This approach aligns with current trends in applied linguistics that emphasize formative assessment, learner autonomy, and the development of higher-order writing skills. Future research should focus on longitudinal studies, cross-linguistic comparisons, and classroom-based applications of coherence-focused AWE systems to further validate and refine these approaches.

RESEARCH METHODOLOGY

This chapter outlines the methodological procedure used to examine how the DECOR framework can be integrated with the EFCAMDAT learner corpus to support coherence-focused assessment in second language (L2) writing. The methodology is divided into the following sections: research design, data source, sampling procedures, annotation method, NLP application, evaluation criteria, ethical considerations, and methodological limitations.

3.1 Research Design

The study follows a mixed-methods research design that combines quantitative analysis with qualitative evaluation. The quantitative aspect focuses on calculating NLP-based coherence indicators and conducting statistical comparisons between learner texts before and after revision through the DECOR framework. This allows the study to measure the effectiveness of automated coherence improvement in a systematic way.

The qualitative aspect involves expert human raters who assess selected learner texts to determine whether the revisions generated by DECOR are pedagogically meaningful and consistent with human expectations of coherent writing. This combined design ensures both large-scale measurement and deeper interpretive understanding of writing quality and learner development.

3.2 Data Source

The main source of data for this research is the EF-Cambridge Open Language Database (EFCAMDAT), a publicly available learner corpus containing more than 1.18 million English texts written by second language learners. The corpus covers all CEFR proficiency levels from A1 to C2 and includes important metadata such as learner age, first language background, writing task type, and revision history.

Because of its large scale and detailed learner information, EFCAMDAT provides a strong foundation for investigating coherence development and testing NLP-based writing assessment tools across multiple proficiency levels.

3.3 Data Selection Criteria

To represent learners from intermediate to advanced proficiency levels, a stratified sample of 300 learner texts was selected from four CEFR levels: A2, B1, B2, and C1. Each level contributed 75 texts to ensure balanced representation across proficiency groups.

The selected texts were controlled for both task type and length, with each composition ranging between 150 and 250 words. Only texts with clear paragraph structure and sufficient discourse development were included so that coherence could be meaningfully examined. In addition, learners' first language (L1) background information was recorded to observe possible influences of L1 transfer on coherence patterns.

3.4 Annotation Process

Before applying the DECOR model, a subset of 100 learner texts was manually annotated for coherence-related problems by two trained linguists. The annotation guidelines were developed

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

following the framework proposed by Zhang et al. (2024), focusing on issues such as topic drift, weak transitions, semantic inconsistency, and lack of logical progression.

To ensure reliability in annotation, inter-annotator agreement was calculated using Cohen's Kappa coefficient. These annotations served both as a benchmark for evaluating DECOR's detection accuracy and as supplementary training data for further model refinement where necessary.

3.5 NLP Implementation

The DECOR framework (Detect, Explain, and Rewrite) was implemented using the pre-trained benchmark model introduced by Zhang et al. (2024). The framework operates through three consecutive stages:

• **Incoherence Detection:** locating sentences or segments where logical flow is disrupted.
• **Reasoning:** identifying the specific cause of incoherence, such as abrupt transition, semantic mismatch, or unclear connection between ideas.
• **Rewriting:** generating an improved version of the problematic segment to strengthen coherence and readability.

This pipeline was applied to all selected learner texts. The original and revised versions were then compared using both automated tools and human evaluations to determine the effectiveness of the DECOR framework.

3.6 Evaluation Metrics

The effectiveness of DECOR was assessed through a combination of automated and human-based evaluation methods.

• **Automated Metrics:** Coh-Metrix and TAACO were used to measure changes in cohesion-related features such as connectives, lexical overlap, referential cohesion, and semantic similarity.
• **Human Ratings:** Three experienced ESL instructors evaluated both original and revised drafts using a 5-point coherence rating scale adapted from IELTS Writing Band Descriptors. Their assessments focused on idea progression, logical organization, paragraph unity, and clarity of transitions.
• **Statistical Analysis:** Paired t-tests and ANOVA were conducted to examine whether the differences in coherence scores before and after DECOR intervention were statistically significant across CEFR proficiency levels.

3.7 Ethical Considerations

The study uses data from EFCAMDAT, which is publicly accessible and fully anonymized for research purposes. No personal or identifiable learner information is included in the analysis, ensuring compliance with ethical research standards.

Human raters involved in the evaluation process participated voluntarily and were provided with informed consent forms explaining the purpose of the study, their responsibilities, and their right to withdraw at any stage. All collected data were securely stored and used strictly for academic research purposes only.

3.8 Limitations of the Methodology

Although DECOR offers an innovative approach to improving coherence, its rewriting suggestions may not always fully match pedagogical expectations, particularly in creative writing tasks or genre-specific compositions where multiple interpretations may be acceptable.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

In addition, while EFCAMDAT provides valuable learner metadata, it does not fully reflect classroom teaching conditions, teacher feedback practices, or learners' instructional experiences, all of which may influence writing development. The exclusive focus on English learner texts also limits the broader applicability of the findings to other target languages.

Future studies may address these limitations by incorporating classroom-based research designs and multilingual learner corpora to provide a broader understanding of coherence development across different language learning contexts.

RESULTS AND FINDINGS

This section presents the findings of the study by showing how the DECOR NLP framework contributes to improving coherence in second language (L2) writing. The results are reported in line with the methodological procedures outlined earlier and are based on both automated NLP measurements and human evaluator judgments. The analysis focuses on several key areas: changes in coherence scores before and after DECOR intervention, the frequency and categories of coherence problems identified, the degree of consistency between machine-generated assessments and human ratings, and qualitative observations drawn from selected sample texts. To make the findings clearer and more systematic, four tables are included in this section.

4.1 Improvement in Coherence Scores (Automated Metrics)

To examine the effect of DECOR on textual coherence, pre-intervention and post-intervention scores were generated using Coh-Metrix and TAACO, two widely used NLP tools for analyzing cohesion and coherence. These tools evaluate several discourse-related features, including lexical overlap, referential cohesion, use of connectives, and Latent Semantic Analysis (LSA) similarity. The comparison of scores across CEFR levels provides an overview of how coherence changed after DECOR-based revision. Table 1 presents the results of these automated analyses.

Table 1

Pre- and Post-Intervention Coherence Scores Across CEFR Levels

CEFR Level	Coh-Metrix LSA Cohesion (Pre)	Coh-Metrix LSA Cohesion (Post)	TAACO Lexical Overlap (Pre)	TAACO Lexical Overlap (Post)
A2	0.45 (±0.05)	0.55 (±0.04)	0.30 (±0.06)	0.42 (±0.05)
B1	0.52 (±0.04)	0.63 (±0.05)	0.38 (±0.05)	0.47 (±0.04)
B2	0.60 (±0.03)	0.68 (±0.04)	0.45 (±0.04)	0.53 (±0.03)
C1	0.66 (±0.02)	0.72 (±0.03)	0.50 (±0.03)	0.57 (±0.02)

The results show a clear improvement in coherence across all CEFR levels after DECOR-based intervention. Both Coh-Metrix LSA cohesion scores and TAACO lexical overlap scores increased from pre-test to post-test, suggesting that the revised texts demonstrated stronger logical flow and better lexical connectedness.

4.2. Error Typology and Frequency Analysis

To examine the coherence problems targeted by DECOR, every sentence identified and revised by the system was classified according to the framework's error categories. The resulting distribution of these coherence-related issues is presented in Table 2.

Table 2

Frequency of Coherence Error Types Identified by DECOR

Error Type	Frequency	Percentage (%)
Semantic Mismatch	220	40.0

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

Topic Drift	180	32.7
Abrupt Transition	150	27.3
Total	550	100.0

The analysis shows that semantic mismatch was the most frequent coherence issue, followed by topic drift and abrupt transition. These findings indicate that learners across proficiency levels continue to face challenges in maintaining logical consistency and smooth progression of ideas in writing. The analysis indicates that semantic mismatches occurred most frequently in lower and mid-level learner texts, where ideas were often inconsistent or did not logically connect with surrounding content. Topic drift appeared more often in texts at the B1 and B2 levels, suggesting difficulty in maintaining a clear central focus across a response. Abrupt transitions were also observed in higher-level writing, showing that even more proficient learners may continue to struggle with smooth discourse organization and logical sequencing.

4.3. Coherence Gains by Feature Category

To obtain a clearer picture of where improvement occurred, changes in coherence were examined across specific discourse features, including connectives, referential cohesion, and lexical chains. Mean score differences before and after the intervention were calculated from 120 representative essays, and the results are presented in Table 4.

The findings show that the strongest improvements appeared in entity continuity and referential cohesion. This suggests that DECOR was particularly effective in improving clarity of reference and sustaining topic consistency throughout the text, both of which are essential components of coherent L2 writing.

Table 3

Coherence Gains by Feature Category

Feature Category	Pre-Intervention Mean	Post-Intervention Mean	% Gain
Connective Density (/sent)	1.42	2.04	43%
Referential Cohesion	0.28	0.41	46%
Lexical Overlap	0.35	0.49	40%
Entity Continuity (TAACO)	0.22	0.35	59%

The results indicate that the greatest improvement occurred in entity continuity and referential cohesion. This suggests that DECOR effectively strengthened clarity of reference and maintained topic consistency, which are important elements of coherence in second language writing.

5. Qualitative Examples and Pedagogical Relevance

The qualitative analysis showed that DECOR generated revisions that were sensitive to context and generally consistent with the expectations of academic writing. For example, in one B1-level learner text, the original sentence was:

“I think exercise is good but also not because sometimes people not want to do it.”

DECOR revised it as:

“Although exercise is beneficial, some people may lack motivation to engage in it.”

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

This revised version demonstrates several improvements. It restructures the sentence more effectively, uses a clearer concessive connector ("although"), and presents the idea in a tone more suitable for academic discourse. On average, human evaluators rated the revised sentence 1.5 points higher on a 5-point coherence scale.

Additional qualitative improvements were also observed, including:

• better maintenance of topic consistency across paragraphs.

• clearer referential expressions, especially where vague pronouns such as "this" or "it" were replaced with more specific nouns.

• stronger logical connections through the use of appropriate transition markers, for example replacing informal or disconnected wording such as "also good" with a more coherent expression like "in addition."

These revisions directly support important pedagogical aims by helping L2 learners produce writing that is more organized, explicit, and logically connected. Taken together, the findings suggest that DECOR contributes meaningfully to improvements in NLP-based coherence measures across CEFR levels. The framework was effective in identifying and revising common discourse-related problems, including semantic mismatch, topic drift, and weak transitions. Human raters also showed broad agreement with the coherence improvements generated by the system. In addition, feature-level gains in areas such as referential cohesion and lexical overlap offered more detailed evidence of how coherence was strengthened. Overall, the revised outputs were contextually appropriate and broadly consistent with academic writing conventions. These results further support the value of NLP in L2 writing assessment and instruction, and they indicate that DECOR has potential not only as an assessment mechanism but also as a pedagogical resource in language learning environments.

5. DISCUSSION

The findings of this study demonstrate that the DECOR NLP framework has strong potential as a coherence-focused tool for second language (L2) writing assessment and feedback. Through the combination of quantitative results and qualitative evaluation, the study shows that DECOR is capable of identifying discourse-level weaknesses in learner writing and generating revisions that are contextually appropriate and closely aligned with human judgment. This chapter discusses these findings in relation to previous research, examines their implications for L2 writing pedagogy and assessment, addresses the limitations of the study, and offers suggestions for future research.

5.1 Coherence Improvements

Coherence in writing depends largely on the logical connection and consistency of ideas throughout a text. When ideas are clearly linked and the progression of thought is easy to follow, the writing is considered coherent. The improvements shown in Coh-Metrix and TAACO scores after DECOR intervention provide strong evidence that computational tools can make measurable contributions to writing quality.

The increase in LSA-based semantic similarity and lexical overlap is especially important because these features are strong indicators of perceived coherence (Hartwell & Aull, 2021). The greater improvement observed among lower CEFR learners suggests that beginner and intermediate writers benefit most from explicit support in cohesion. This supports earlier findings by Hartwell and Aull, who argued that L2 learners often underuse cohesive devices because of limited lexical knowledge and restricted control of sentence structure.

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

DECOR helped address this problem by inserting contextually appropriate cohesive devices, improving lexical links, clarifying references and pronouns, and strengthening clause relationships through discourse markers such as “although,” “because,” and “as a result.” The particularly high gains in referential cohesion and entity continuity indicate that the framework is especially effective in resolving pronoun ambiguity and maintaining topic consistency, both of which are major challenges in L2 writing.

5.2 Observations from Error Typology Analysis

The frequency analysis of coherence errors revealed that semantic mismatch and topic drift were the most common problems across learner texts. This reflects long-standing difficulties in L2 composition, particularly in discourse planning and idea organization. Semantic mismatches occur when learner sentences contradict surrounding ideas or fail to connect logically with the broader context, while topic drift weakens coherence by causing learners to move away from the central theme of the text.

Topic drift may result from limited genre awareness or weak organizational strategies, both of which reduce the clarity expected in academic writing. The ability of DECOR to identify and revise these issues is significant because it provides the type of focused and individualized feedback that is often only possible in one-to-one instruction.

From a pedagogical perspective, this is highly valuable, especially in large classrooms where teachers may not have enough time to provide detailed discourse-level feedback to every learner. DECOR therefore functions not only as an assessment tool but also as a practical support system for writing development.

5.3 Alignment with Human Evaluation

One of the strongest findings of the study is the high level of agreement between DECOR-generated coherence scores and human evaluator judgments. The strong relationship between machine ratings and human ratings confirms the validity of DECOR as a reliable assessment tool. A Pearson correlation coefficient of 0.75 and a Cohen’s Kappa value of 0.70 indicate that the system’s judgments are not only computationally accurate but also pedagogically meaningful.

This finding supports earlier research such as Zhang et al. (2024), which showed that well-trained neural models can closely approximate expert human evaluation in writing assessment. At the same time, small differences between system ratings and human judgments in higher-level texts suggest that DECOR may still struggle with more subtle coherence strategies used by advanced writers, such as inferential transitions, implicit topic shifts, and sophisticated rhetorical development.

This indicates that while DECOR performs strongly in identifying visible discourse problems, further refinement may be needed for evaluating advanced academic writing.

5.4 Feature-Level Analysis and Instructional Implications

The feature-based analysis presented in Table 3 provides deeper insight into how DECOR contributes to writing improvement. The 59% increase in entity continuity and the 46% improvement in referential cohesion show that the system does more than simply detect discourse entities—it also strengthens the consistency of how those entities are maintained across the text.

From an instructional perspective, these results suggest that DECOR can be highly effective in supporting the explicit teaching of coherence and cohesion. If the system provides not only

ISSN E: 2709-8273
ISSN P:2709-8265
JOURNAL OF APPLIED LINGUISTICS AND TESOL
JALT

revised versions but also explanations for the revisions, it can become an even stronger teaching resource.

This aligns with principles of Data-Driven Learning (DDL), where learners improve their writing by observing authentic language patterns and understanding how errors are corrected rather than simply memorizing abstract rules (Schmidt, 2022). In this way, DECOR can support corpus-informed writing instruction and help learners develop stronger discourse awareness.

5.5 Pedagogical Value of Qualitative Rewrites

The qualitative examples presented earlier illustrate how DECOR can transform learner writing into forms that are more consistent with academic writing standards. By rewriting unclear or disconnected sentences, the system was able to introduce better subordination, reduce unnecessary repetition, and restore logical connections between ideas.

These are central features of effective academic writing and are often difficult for L2 learners to master independently. This suggests that DECOR can function not only as an assessment mechanism but also as a digital instructional assistant that models appropriate academic language use.

Such support is particularly useful in educational environments where teachers have limited time for individual writing conferences. In these situations, DECOR can provide immediate revision suggestions that complement teacher feedback rather than replacing it.

5.6 Integration into Formative Assessment Practices

The use of DECOR in formative assessment contexts offers strong practical potential. Since writing development often involves multiple drafts and continuous revision, DECOR can be used to provide immediate coherence-focused feedback during the drafting process. This encourages learners to revise their writing based on discourse quality rather than focusing only on grammar correction.

This approach is consistent with best practices in formative assessment, where timely and constructive feedback plays a major role in improving learning outcomes (Ariely et al., 2020). In addition, interacting with system-generated revisions may help learners develop stronger metalinguistic awareness and self-editing skills.

By comparing their own writing with improved versions produced by DECOR, learners can become more aware of coherence problems and gradually develop greater independence in managing discourse structure. This supports learner autonomy, which remains one of the central goals of second language writing instruction.

6.CONCLUSION

This study examined the use of the DECOR (Detect, Explain, and Rewrite) NLP framework in second language (L2) writing assessment, with particular emphasis on coherence as one of the most essential yet often under-evaluated components of effective writing. Using learner texts from the EF-Cambridge Open Language Database (EFCAMDAT), the research demonstrated that coherence-focused NLP tools can significantly contribute to writing improvement, strengthen instructional practices, and promote greater learner engagement in the writing process. Both quantitative findings from NLP-based metrics and qualitative analysis of revised learner texts confirmed the strong potential of DECOR as a tool for formative assessment and instructional support. One of the major contributions of this research is the empirical evidence showing measurable improvement in coherence through automated rewriting. The use of Coh-Metrix and TAACO revealed clear increases in semantic similarity and lexical cohesion,

particularly among learners at lower CEFR levels such as A2 and B1, where coherence-related difficulties are more frequent. Important discourse features including referential continuity, entity overlap, and the use of connectives showed noticeable improvement after DECOR intervention. These changes were supported not only by computational analysis but also by positive human evaluator judgments, which strengthened the pedagogical validity of the framework. In addition to improving textual quality, DECOR proved effective in identifying and correcting common coherence problems such as topic drift, semantic inconsistency, and abrupt transitions. The system produced revisions that were context-sensitive and closely resembled the kind of focused feedback typically provided by experienced writing instructors. This makes the framework particularly valuable for learner scaffolding and writing development. From a theoretical perspective, the study also contributed by transforming the traditionally subjective concept of coherence into identifiable error categories and measurable NLP indicators. This supports the growing movement in applied linguistics toward multidimensional writing assessment that includes discourse-level evaluation.

The educational and technological implications of the findings are significant. DECOR can be integrated into digital learning platforms to provide immediate feedback during drafting or function as a post-writing assessment tool. Because of its flexible structure, the framework can be adapted for learners at different proficiency levels by adjusting the focus on detection, explanation, and guided rewriting. The findings also suggest that DECOR may have value in summative assessment settings where broader dimensions of writing quality need to be considered. Such applications show strong potential for supporting personalized learning as well as large-scale educational contexts. Despite these strengths, the study also identified several limitations. Although many of DECOR's rewrites were effective, some revisions appeared overly mechanical, simplified meaning too much, or reduced the naturalness of idiomatic and creative language use. In some cases, the explanation component also lacked sufficient clarity for direct pedagogical use. Ethical concerns related to AI in education must also be considered, including issues of algorithmic bias, data privacy, and excessive dependence on automated systems. Future research should explore cross-linguistic applications, learner interaction with system feedback, and the long-term effects of DECOR on writing development. Overall, this study helps bridge the gap between computational linguistics and classroom pedagogy, positioning DECOR as a valuable resource in the future of language learning and writing assessment.

REFERENCES

Abraha, T., & Nazir, A. (2024). *Evaluation of transformer-based neural language models for writing feedback and automated essay scoring*. https://doi.org/10.21203/rs.3.rs-3979085/v1

Ariely, M., Nazaretsky, T., & Alexandron, G. (2020). *First steps towards NLP-based formative feedback to improve scientific writing in Hebrew*. https://doi.org/10.35542/osf.io/pe5ky

Delu, Z. (2021). Cohesion in multimodal text. In *New research on cohesion and coherence in linguistics* (pp. 182–201). https://doi.org/10.4324/9781003190110-10-13

Granger, S. (2021). Phraseology, corpora and L2 research. In *Perspectives on the L2 phrasicon* (pp. 3–22). https://doi.org/10.21832/9781788924863-002

Hanaoka, O., & Izumi, S. (2021). Directions for future research on attention and L2 writing. In *The Routledge handbook of second language acquisition and writing* (pp. 312–324). https://doi.org/10.4324/9780429199691-32

Hartwell, K., & Aull, L. (2021). Automated text-matching and writing-assistance tools. *Assessing Writing, 50*, Article 100562. https://doi.org/10.1016/j.asw.2021.100562

Koopman, H., & Guardiano, C. (2022). Managing data in TerraLing, a large-scale cross-linguistic database of morphological, syntactic, and semantic patterns. In *The open handbook of linguistic data management* (pp. 617–630). https://doi.org/10.7551/mitpress/12200.003.0060

Mayormente, M. D., & Gumpal, B. R. (2025). NLP-based sentiment analysis for evaluating student feedback in English language education. In *2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)* (pp. 112–118). https://doi.org/10.1109/ICMSCI62561.2025.10894214

Schmidt, N. (2022). Unpacking second language writing teacher knowledge through corpus-based pedagogy training. *ReCALL, 35*(1), 40–57. https://doi.org/10.1017/S0958344022000106

Zhang, X., Diaz, A., Chen, Z., Wu, Q., Qian, K., Voss, E., & Yu, Z. (2024). DECOR: Improving coherence in L2 English writing with a novel benchmark for incoherence detection, reasoning, and rewriting. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing* (pp. 11436–11458). https://doi.org/10.18653/v1/2024.emnlp-main.639