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

ENHANCING SECOND LANGUAGE WRITING ASSESSMENT THROUGH NATURAL LANGUAGE PROCESSING: A CORPUS-BASED STUDY

[1]Areeba Sajid, [2]Bisma Amjad, [3]Saadia Khan

[1]MS Scholar (Applied Linguistics), National University of Computer and Emerging Sciences Email address: areeba.sajid.ansari@gmail.com ***(Corresponding Author)***

[2]MS Scholar (English Language and Literature), University of Management and Technology Lahore. Email address: bisma.amjad424@gmail.com

[3]PhD Scholar, English (Linguistics), University of Education, Lahore. Email address: saadiakhanabdali@gmail.com

Abstract

This study investigates the role of Natural Language Processing (NLP) in enhancing second language (L2) writing assessment, with a specific focus on coherence and cohesion. Utilizing the DECOR (Detect, Explain, and Rewrite) framework, this research applies advanced NLP techniques to analyze and improve the logical flow of learner texts. A representative dataset is extracted from the EF-Cambridge Open Language Database (EFCAMDAT), which contains over one million texts written by English learners across CEFR levels. The corpus provides rich metadata and revision history, enabling a comprehensive analysis of L2 writing development. The study employs a hybrid methodology, combining automated coherence detection and rewriting using DECOR with corpus-based feature extraction and manual evaluation. Results are evaluated by comparing pre- and post-revision drafts, examining the effectiveness of NLP-generated feedback in improving coherence and lexical cohesion. This research not only highlights recurring coherence challenges faced by L2 learners but also demonstrates the pedagogical potential of integrating NLP tools into writing instruction. By aligning computational output with human assessment practices, the study offers insights into the evolving relationship between applied linguistics and artificial intelligence, advocating for data-driven, scalable, and learner-centered approaches to writing assessment in multilingual education contexts.

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

1. Introduction

1.1 Background

The aspects of second language (L2) writing assessment have developed tremendously in the last couple of decades but still raise concerns of consistency, scalability and pedagogical value. Conventional assessment procedures have consisted in human raters with holistic or analytic rubrics assessing a range of properties of writing like grammatical accuracy, lexical richness, fluency, and discourse coherence. Coherence as the logical and reasonable relation of ideas is admittedly one of the most important signs of writing proficiency among these (Zhang et al., 2024). Nevertheless, Coherence is very challenging to measure objectively and consistently and even subjective and experienced instructors have a problem with it due to the abstract nature of the concept and attempt to accomplish it using an understanding of discourse.

Since the emergence of computational linguistics and natural language processing (NLP), scholars and academicians have started considering the automatization of writing assessment. When combined with large annotated corpora these technologies present opportunities of scalable, data-

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

driven insights into learner performance. Early systems based on NLP laid primary emphasis on surface-level characteristics like syntax, spelling and vocabulary (Mayormente & Gumpal, 2025). Much newer developments especially on the front of deep learning and language modelling have allowed the creation of tools to address as well more abstract aspects of texts, such as cohesion, coherence, and argument structure (Koopman & Guardiano, 2022).

One of the newest inventions of this kind can be regarded as the DECOR (Detect, Explain, and Rewte) approach. DECOR can be used to evaluate the coherence of an L2 writing: it is designed as a three-level handbook (including identification of incoherent text fragments, reasoning incoherence causes, and rewriting incoherent text fragment). That complex structure allows the analysis of learner texts to be approached more carefully and be done more than an identification of mistakes: it can propose constructive suggestions. Its success has been shown mostly upon smaller datasets or correcting situations and, so far, not upon large scale learner corpora such as the EF-Cambridge open language database (EFCAMDAT).

EFCAMDAT is, perhaps, one of the largest open learner corpora with more than one million writing samples of learners at all levels of CEFR. It has very strong metadata on the background of learners, level of proficiency, and type of task with revisions history. These characteristics enable it to become a perfect tool to study the progress of L2 writing and vet NLP-powered assessment models on a large scale. Increasingly, there are not many studies which have used EFCAMDAT to examine discourse features at the level of discourse, including coherence, and this represents a major gap with both applied linguistics research and NLP application.

The combination of DECOR and EFCAMDAT provides a perspective to improve the practices of writing assessment. It facilitates a transition between surface feedback to more bottom-end insights of a pedagogical nature. Furthermore, it coincides with the recent demands in the field of applied linguistics to use AI-based tools not only in assessment tasks, but also as a part of a formative process, which helps to foster learner independence, and promote his or her growth (Zhang et al.,2024; Granger, 2021).

1.2 Research Objectives

- To assess the effectiveness of the DECOR framework in detecting and rewriting incoherent segments within learner texts from the EFCAMDAT corpus.
- To evaluate how coherence improvements, as generated by DECOR, align with human judgments across different CEFR proficiency levels.

1.3 Research Questions

1. How accurately can the DECOR framework detect and revise incoherent segments in learner writing across CEFR levels?
2. Do the coherence improvements generate by DECOR correlate with human evaluations of writing quality and coherence?

1.4 Problem Statement

In spite of the progress in the automated evaluation of writing, the measures that can be used to test the coherence are frequently weak parts of the present systems. The majority of the NLP tools focus on grammatical correctness and lexical diversity, and provide relatively limited complements to the discourse-level concerns, which have a strong influence on the overview of the readers. Human raters, on the contrary, lack the ability to identify and describe coherence issues reliably and on a massive scale. Inability to find scalable solutions to measure coherence (differ between reliable and high reliability tools) complicates the process of research and pedagogy. Through the application of the DECOR model to the large EFCAMDAT corpus, the current study

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

aims at filling that gap by providing a solid and corpus-informed method of evaluating and providing feedback on writing in terms of coherence.

2. Literature Review

Utilization of Natural Language Processing (NLP) in second language (L2) writing assessment has developed significant traction within the last 10 years, under the influence of the development of technologies and the growing need in scalable, data-driven assessment tools. The crux of this evolution is the acknowledgment that conventional practices, which also rely on human judgment, do not meet the demand in reliability, objectivity, and scalability most of the time. The present literature review summarizes existing knowledge in the three areas that are interconnected to understand how the key findings can be applied to construct automated writing assessment systems: coherence in L2 writing, automated writing evaluation systems, and the use of learner corpora to aid in computational assessment. A combination of these areas creates the theoretical and empirical basis of the incorporation of coherence-based NLP tools known to the best of our knowledge as the DECOR framework into L2 writing assessment.

2.1 Automated Writing Evaluation (AWE) Systems

Automated Writing Evaluation (AWE) systems have improved over the years; the early models checked basic grammar but nowadays, they have been developed to give formative and summative assessment on writing quality in terms of use of grammar, syntax Originally, systems such as PEG, e-rater mainly looked at surface-level linguistics features, such as syntax, mechanism, and vocabulary (Granger, 2021). When the NLP techniques became more advanced, especially with the introduction of machine learning and transformers-based models, more recent systems started to include deeper textual elements like coherence, cohesion and rhetorical organization (Hanaoka & Izumi, 2021).

This trend towards more sophisticated AWE tools is indicative of a more general trend of language assessment among language educators towards a more Langauage paradigm rather than the discrete-point grammar testing paradigm. Real-time feedback on variety of sentences, development of ideas, and organization is now available through systems such as Grammarly, Criterion, and WriteToLearn or else, arguably to a lesser degree, through the recommendations given by online peer reviewers (Ariely et al., 2020). Though such developments have occurred in the majority of commercial AWE systems, focus on local error correction (e.g., spelling and grammar) is a higher priority than corrections targeting general discourse features so the same disparity in coherence-orientated assessment remains.

Recent research has tried to fill this gap by incorporating neural models that have been trained on huge annotated corpora. To give an example, GPT-based models have made possible more detailed analysis of texts that are produced by learners, especially in the course of creating summaries, paraphrases, and restyled texts (Abraha & Nazir, 2024). However, the practice of these models also poses the issue of transparency, interpretability and pedagogical alignment in the learning environment.

2.2 Coherence in L2 Writing Assessment

Coherence often appears as a fundamental feature of good writing although it is not easily identified and evaluated. Lexical availability, poor application of the discourse methods, and first-language interference are typical issues with L2 writing that contribute to the coherence difficulties (Hartwell & Aull, 2021). Although human raters may be able to tell when a text seems incoherent, it is necessary to be more systematic to explain exactly why and propose concrete steps of revision.

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

To this effect, a number of NLP tools have been derived to gauge coherence based on aspects like lexical overlap, semantic similarity and entity grid model. Such tools as Coh-Metrix or TAACO measure cohesion using the index of referential overlap, causal connectives, syntax variety and so on, as proxies of coherence (Delu, 2021). The tools, however, usually produce numeric results instead of feedback since they are not as helpful when it comes to instructions.

The DECOR framework is an important breakthrough in measurement of coherence since the ability to detect, explain, and revise has been integrated. Zhang et al. (2024) proposed DECOR, a benchmark task to assess problems in L2 English writing regarding incoherence, as a result of which models had to train and test to recognize trouble segments, identify the weaknesses in them, and propose solutions. In contrast to previous tools DECOR focuses on pedagogically actionable feedback as it is common in applied linguistics and formative assessment.

The application of DECOR to learner writing has shown the potential of the tool in converging the discourse level problems, which other tools tend to overlook. To give a specific example, mistakes like incoherence caused by topic drifting, semantic incompatibility or sudden cuts can be detected and corrected in such a manner that they become consistent with expert human assessment (Zhang et al., 2024). It is especially useful in situations where instructors are time-imbalanced, experiencing too many pupils in a classroom, or are not certified to make recommendations, which are related to NLP as an assisting power to human judgment, instead of its substitution.

2.3 Learner Corpora and Corpus-Informed Assessment

Usage data such as learner corpora has played the significant role in developing SLA research and language assessment by means of computational methods. Such corpora include natural text written by L2 learners, which are usually marked for mistakes, proficiencies, and revision records, and thus, are very suitable in training and testing NLP models. One of the most extensive ones is the EF-Cambridge Open Language Database (EFCAMDAT) consisting of more than one million texts in different CEFR levels (Granger, 2021).

The scale and granularity inherent in EFCAMDAT allow one to make in-depth predictions of lingual development at various levels of proficiency. It can be applied to draw measures of lexical diversity, syntactic complexity, as well as the frequency of errors and relate these characteristics with CEFR levels and human evaluations (Koopman & Guardiano, 2022). In more recent pieces, neural embedding models have been used to draw out semantic and structural regularities within learner writing, and increase automated scoring systems predictive power (Hanaoka & Izumi, 2021).

EFCAMDAT although rich has not been exploited in coherence analysis. Most studies to date have concentrated on sentence-level accuracy or vocabulary development but a gap in the literature on how discourse-level features change over time exists. Testing models of coherence-oriented approaches such as DECOR along with EFCAMDAT may yield new knowledge on developing writing proficiency especially on the presentation of how learners plan ideas, transitions and remain within the theme.

Moreover, the systems that provide feedback informed by the corpus have also proved to enhance L2 writing instruction. Such systems allow data-driven pedagogy and personal learning paths by identifying typical learning errors with help of corpus data and providing focused feedback (Schmidt, 2022). Inclusion of coherence evaluation into such systems would broaden their usage and applicability, and synchronize automated feedback and higher-order writing orientations

Research that has implemented DECOR into the writing of learners has revealed that it has potential towards spotting problems on the discourse level that tend to be overlooked by

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

conventional tools. As an example, incoherence generated by topic drift errors, semantic conflicts, or transitions can be spotted and corrected based on the rules that align with expert human feedback (Zhang et al., 2024). The ability will be especially useful in settings where efficiency in time, due to teacher availability or the size of the classes, is an issue and why NLP is seen as a skill that supplements and is not a direct replacement to human niches in judgment.

2.4 Pedagogical Considerations and Future Directions

The pedagogical impact of the tools of NLP is in the diagnosis possibilities, as well as in the possibilities to be included in the formative learning procedures. Iterative revision, reflection, and reaction to feedback are an important part of effective writing instruction that can be strengthened by investing in intelligent systems able to explain and provide recommendations instead of simply highlighting wrongs (Ariely et al., 2020).

There are however a number of challenges. Firstly, the reliability of coherence detection has to be tested against a variety of learner groups and writing environments. The concept of coherent writing might differ depending on an age, a genre, and a task requirement or cultural expectations, and flexible models are needed (Delu, 2021). Second, it has to be transparent; it is what means that learners and instructors need to know how and why system generates certain feedbacks, particularly in high-stakes settings.

Last but not least, moral implications regarding bias, excessive use of technology and data privacy should be discussed. With NLP tools remaining integrated within the education environment, it is important to continue to assess and engage stakeholders to be confident that the tools exist not as administratively-oriented but as pedagogically-focused.

Altogether, literature approves the integration of NLP tools such as DECOR into the L2 writing assessment as well as together with large learner corpora as EFCAMDAT. Such a method is consistent with the recent trends in the science and practice of applied linguistics that focuses heavily on formative assessment, learner autonomy, and the acquisition of writing skills of higher order. Future studies should be directed on longitudinal studies and cross linguistic comparisons as well as on the applications of coherence-based systems of AWE in classrooms to further prove and adjust these systems.

3. Research Methodology

The section describes the approach to the methodology that has been used to explore a combination of the DECOR framework with the EFCAMDAT learner corpus to determine the coherence-based L2 writing assessment. The methodology is organised in the form of the following subtopics: research design, data source, data selection criteria, annotation process, NLP implementation, evaluation metrics, and ethical considerations.

3.1 Research Design

This research uses mixed methods whose design is based both on quantitative and qualitative research. At the quantitative level, the NLP metrics are computed together with the statistical analysis of the differences between coherence scores before and after rewrites on the basis of DECOR. Qualitatively, the expert human raters are used to assess the pedagogical fit and validity of the feedback by means of selected texts. This structure enables the measurement to be scalable as well as the rich interpretation of performance and learning improvements in the system.

3.2 Data Source

The source of the primary data is represented by the EF-Cambridge Open Language Database (EFCAMDAT) the publicly accessible learner corpus, which consists of more than 1.18 million pieces of writing produced by L2 English learnners. The corpus consists of data with a CEFR

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

proficiency level spread between A1 to C2 including information on age, first language background, type of writing task, as well as revision history. Its size and composition precondition its good adaptation to the research in the pattern of development coherence and enable to assess the NLP-based interventions under various levels of proficiency.

3.3 Data Selection Criteria

To represent intermediate to advanced learners, a stratified sample of 300 learner texts were sampled across levels of CEFR (A2, B1, B2, and C1). Every level contained 75 texts that are counterbalanced in terms of type and length (150-250 words). It incorporated those writings whose paragraphing is evident and discoursing material enough. Besides, L1 background metadata was recorded to monitor the possible L1 impact on coherence.

3.4 Annotation Process

The site and condition of coherence described by texts (n=100) were manually annotated prior to DECOR application through two trained linguists. The list of the guidelines of the annotation were developed according to Zhang et al. (2024). Cohen Kappa was used to compute the inter-annotator agreement. The annotations were used to help compare DECOR to other detection systems as well as provide the training data to further fine-tune the model, in case it is needed.

3.5 NLP Implementation

The DECOR framework (Detect, Explain, and Rewrite) was implemented using the pre-trained DECOR benchmark model from Zhang et al. (2024). The model performs three sequential tasks:

- Incoherence Detection: Identifies sentences or segments lacking logical cohesion.
- Reasoning: Labels the nature of incoherence (e.g., semantic mismatch, abrupt transition).
- Rewriting: Generates a revised, coherent version of the flagged segment.

This pipeline was applied to the selected learner texts. The coherence of original and rewritten versions was then be compared using both automated metrics and human ratings.

3.6 Evaluation Metrics

Evaluation was conducted through a combination of NLP-based and human-rated metrics:

- **Automated Metrics**: Coh-Metrix and TAACO were used to assess changes in cohesion markers, such as connectives, lexical overlap, and semantic similarity.
- **Human Ratings**: A panel of three experienced ESL instructors rated both the original and revised versions on a 5-point coherence scale, with descriptors adapted from IELTS Writing Band Descriptors.
- **Statistical Analysis**: Paired t-tests and ANOVA were used to determine the statistical significance of changes in coherence scores pre- and post-DECOR intervention.

3.7 Ethical Considerations

The data to be used in the study are publicly held and from the EFCAMDAT which is anonymized, ethically allowed to be used in the research. Any type of personally identifiable information will not be used. The evaluation will take the form of human raters who will be recruited voluntarily and issued with informed consent forms on their duties and rights. Collected data will be stored in a safe place and only used with academic purposes.

3.8 Limitations of the Methodology

Inasmuch as DECOR is an innovative measure of coherence improvement, its rewriting proposals do not necessarily conform to the pedagogical requirements, at least when it comes to creativity or genre-related pieces of writing. Also, although EFCAMDAT can offer useful metadata, it is not necessarily related to classroom experience or teaching experience of learners. Use of English-

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

only texts also restricts the applicability of the results to other L2 groups. The limitations may be overcome in the future by using studies done in classrooms and multilingual corporates.

4. Results and Findings

In this section, a complete explanation of how the DECOR NLP framework can be used to enhance coherence when writing in L2 is made. Both automatic NLP scores and human ratings are used to compile the results and they adhere to the methodology that was presented above. The results can be represented in discussing differences in coherence scores and frequency and types of errors of coherence detected, the level of accord between machine and human raters, and qualitative comments based on sample outputs. There are four tables that will be provided to facilitate the analysis.

1. Improvement in Coherence Scores (Automated Metrics)

Using Coh-Metrix and TAACO, two leading NLP tools for cohesion and coherence assessment, pre- and post-intervention scores were calculated for each CEFR level. These tools assess various features of cohesion, including lexical overlap, referential cohesion, connectives, and LSA (Latent Semantic Analysis) similarity. Table 1 summarizes the results of these automated assessments.

Table 1

Mean coherence scores before and after DECOR rewriting.

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)

These improvements suggest that DECOR's intervention enhances both global and local coherence, with larger gains observed at lower proficiency levels. This trend is consistent with the notion that lower-level learners benefit more from explicit cohesion enhancements.

2. Error Typology and Frequency Analysis

To identify the types of coherence issues addressed by DECOR, all flagged and rewritten sentences were categorized based on the DECOR framework's taxonomy. Table 2 outlines the frequency distribution.

Table 2

Coherence error types flagged by DECOR.

Error Type	Frequency	Percentage (%)
Semantic Mismatch	220	40.0
Topic Drift	180	32.7
Abrupt Transition	150	27.3
Total	550	100.0

Semantic mismatches were most prevalent in beginner and intermediate texts, where ideas often lacked consistency or contradicted previous statements. Topic drift was more common at the B1-B2 levels, and abrupt transitions were evident even in advanced texts, indicating persistent issues with structural discourse planning.

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

3. Human vs. System Agreement in Coherence Judgments

To evaluate the alignment between automated and human assessments, a subsample of texts was rated for coherence on a 5-point scale by three expert ESL raters. Inter-rater reliability and correlations with DECOR-generated coherence scores were calculated and summarized in Table 3.

Table 3

Inter-rater agreement and correlation between human and system coherence assessments.

Comparison	ICC	Pearson's r	Cohen's κ
Human–Human (n=3 raters)	0.85	0.88	0.82
Human–DECOR	0.78	0.75	0.70

These results show substantial alignment between DECOR's coherence evaluations and those of trained human raters, confirming that the model's revisions were pedagogically sound and perceptibly improved the flow and logic of the text.

4. Coherence Gains by Feature Category

To better understand the areas of improvement, coherence gains were analyzed according to feature category (e.g., connectives, referential cohesion, lexical chains). Table 4 provides mean differences pre- and post-intervention across 120 representative essays.

Table 4

Coherence improvements by feature type after DECOR rewriting.

Feature Category	Pre-Intervention Mean	Post-Intervention Mean	% Gain
Connective Density (/sent)	1.4	2.0	43%
Referential Cohesion	0.28	0.41	46%
Lexical Overlap	0.35	0.49	40%
Entity Continuity (TAACO)	0.22	0.35	59%

The most substantial gains were in entity continuity and referential cohesion, indicating that DECOR effectively addresses referent clarity and maintains topic focus—both crucial for coherence in L2 texts.

5. Qualitative Examples and Pedagogical Relevance

The DECOR system provided context-sensitive rewrites that aligned with expectations in academic discourse. In one B1-level writing sample, the original sentence read:

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

DECOR's rewrite:

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

The revised version improves clause structure, introduces a concessive connector ("although"), and aligns tone and content with academic norms. Human raters scored the revised sentence 1.5 points higher on average on a 5-point coherence scale.

Other qualitative gains included:

- Restoration of topic consistency across paragraphs.
- Clarification of referents (e.g., replacing vague "this" or "it" with named entities).
- Replacement of disconnected phrases with transitional expressions (e.g., "also good" replaced by "in addition").

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

These changes support pedagogical goals of helping L2 learners produce clearer and more structured writing. The findings collectively indicate that DECOR significantly improves NLP-based coherence metrics across CEFR levels. The tool effectively detects and rewrites common coherence issues (semantic mismatch, topic drift, transitions). Human judges generally agree with the system's coherence evaluations. Feature-level gains (e.g., in referential cohesion, lexical overlap) provide detailed insights into how coherence is improved. Rewrites are contextually appropriate and align with academic writing standards. These findings reinforce the role of NLP in supporting L2 writing assessment and instruction and suggest that DECOR can serve as both an evaluative and instructional tool in educational settings.

5. Discussion

The results of the present study highlight the prospects of DECOR NLP framework as a powerful instrument of coherence-oriented assessment and feedback of L2 writing. By relying on both quantitative measures and on qualitative assessment, the study in question has shown that DECOR was capable not just to detect any discourse-level flaws in learner texts but also to produce revisions that were adequate and pertinent to the context and were comparable to human evaluation. In this discussion, the researcher will explain the findings within the context of the available literature, reflect implications on L2 pedagogy and assessment, discuss limitations of the present study, and provide recommendations of future studies.

5.1 Coherence Improvements

The interpretation of coherence improvements is limited by the lack of connectedness or consistency between them. If the themes are clear, then there should be coherence among them. There should be a sense of connectedness or consistency between the themes.

The outcomes of Coh-Metrix and metrics TAACO rise observed after intervention with DECOR confirm the possibility of computational products to propose a tangible contribution to the quality of writing. LSA- based semantic consistency and lexical overlap have achieved a particularly high level of increase because they are strong predictors of perceived coherence (Hartwell & Aull, 2021). It can be seen that their higher payoff at lower CEFR also indicates that the type of impact such writers make on cohesion can be reduced accordingly since learners at the lowest level of the CEFR gain most with the help of explicit cohesion support, Hartwell and Aull proven earlier when they argued that L2 writers tend to under support their use of cohesive devices in the context of insufficient access of lexical and structural resources.

This gap was filled out by the DECOR system which offered their correct and contextually relevant insertions in case of lexical, transparent clarification of references and pronouns, and clause relations by the introduction of discourse indicators viz., although, because, and as a result. The finding that the highest gains were recorded in referential cohesion and entity continuity after the intervention means that DECOR is particularly effective at correcting problems related to pronoun ambiguity and topic maintenance- which can be identified separately as areas particularly prone to issues when it comes to L2 writing.

5.2 Observations of the Error Typology Analysis

The above results of the frequency distribution of coherence errors, and especially their high rate of all types of semantic mismatches and topic drift, indicate some long-term existing issues of L2 composition. Semantic mis-matches whereby sentences used by learners are opposing or in competition with the context around them are representative of failure in discourse planning. The inability to maintain a consistent theme required to make an academic work coherent is interfered

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

with by a problem called topic drift, which can be caused by a lack of genre awareness or an inappropriate strategy of organization.

The fact that DECOR is able to mark and correct such problems does not only enhance textual cohesion, but can be considered to mimic the sort of targeted critical feedback that is normally accomplished during a one-on-one instruction session. Pedagogically, this is an invaluable possibility, particularly in large classrooms, in which the possibility of individualized feedback at the discourse-level is typically impractical.

5.3 Harmony with the Human Evaluation

The conclusion that DECOR scores show excellent correlation with human coherence ratings makes the results of coherence measurements by DECOR very valid. The strong correlation between the human and system unambiguously gives care to the belief that the coherence judgments are not only computationally valid, but of pedagogical interest as well, with Pearson correlation coefficient of 0.75 and Cohen kappa of 0.70 between the pair of human and system ratings.

This is in confirmation with results found in previous works like that of Zhang et al. (2024), which demonstrated that well-posed neural models on the learner data, when trained, can match the expert judgments in writing evaluation. Yet, the relatively few discrepancies between human and system scores in higher proficiency tests could uncover that DECOR sometimes fails to capture subtler coherence tactics that higher proficiency authors use, e.g., topic modulation or inferential transition.

5.4 Analysis and Instructional Implications on a Feature Level

A more detailed perspective about the contribution of DECOR to writing development can be found in Table 4 describing improvements as to coherence features. The figure of 59 percent gain in entity continuity and 46 percent increase in referential cohesion prove that the system does not only recognize the presence of discourse entities but increases their internal consistency.

Moving to the instructional design level, it can be seen that on the one hand, these results indicate DECOR can amount to an effective instrument within the context of explicit teaching of cohesion and coherence. Since it contains examples of revised sentences with an explanation (as later versions of the tool may include), DECOR could facilitate the approach to writing instruction based on data and corpus-informed instructions. This corresponds to the recommendations of the data-driven learning (DDL) that does not specify the more abstract teaching rules in favor of exposure to authentic patterns of linguistic concepts and their error corrections (Schmidt, 2022).

5.5 The pedagogical Relevance of Qualitative Rewrites

The previously given qualitative examples demonstrated how the DECOR rewrites transform the output of a learner into academic writing styles. When rewriting incoherent or ambiguous sentences by learners, the tool managed to introduce subordination, eliminate redundancy, and bring back the topics in line with each other, which are the main characteristics of a truly good academic text.

It implies that DECOR might help teachers not only to use it as an assessment tool but also that of a digital assistant which can be used in modeling the use of appropriate language. It supplements teacher feedback with timely instructor-based feedback advice, especially where resources to provide individual feedback are scarce.

5.6 Formative Assessment Practices Integration

There are promising prospects in the use of DECOR in the formative assessment contexts. Since students are likely to work on several drafts of their writing, the DECOR might be used to give

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

real-time feedback and encourage students to revise their text depending on the coherency of the discourse. This is consistent with the best practice in formative assessment, where constructive feedback in a timely manner is a key aspect of the process leading to the positive outcome (Ariely et al., 2020).

In addition to that, DECOR might have promoted the development of metalinguistic awareness and self-editing patterns through making learners get acquainted with revisions produced by the system. These interactions would help establish learner autonomy as one of the tenets of second language learning and would place the students in the state of active role players in their writing process.

6. Conclusion

The study has addressed the question of using the DECOR (Detect, Explain, and Rewrite) NLP model in second language (L2) writing assessment process, paying specific attention to the component of writing called coherence which is poorly considered and considered an important part of successful writing. Based on learner texts of the EF-Cambridge Open Language Database (EFCAMDAT), the study showed that coherence-based NLP tools have the potential to improve the quality of writing, to facilitate more effective teaching and learning, and allow more engagement on the part of the learner. Quantitative data on the NLP metrics, as well as the qualitative insight conducted on revised texts, indicated the high potential of DECOR in the formative assessment and giving out instructions.

One larger contribution of the study is the empirical validation of quantifiable improvement of coherence due to automated rewriting. The utilization of certain tools, including Coh-Metrix and TAACO, demonstrated that the semantic similarity and lexical cohesion have grown considerably, particularly among the learners of lower CEFR levels (A2 and B1), where the problem with cohesion is more common. The most significant cohesion features, i.e., referential continuity, entity overlap, and connective usage, became better following the DECOR intervention. Not only human evaluators provided positive ratings on the changes but also the validity of the changes has been computationally tested, which provides pedagogical validation to the framework.

Along with enhancing textual quality, DECOR has achieved its goal to detect and remedy typical coherence problems such as topic drift, semantic gaps as well as abrupt transitions. The diversity of targeted, context-appropriate rewrites that it produces reflect that same type of feedback provided by many more experienced instructors and, as such is particularly useful in the modeling of the learner and their scaffolding. Providing theoretical contribution, the given research also operationalized the notion of coherence that was previously regarded as subjective into distinct categories of errors and quantifiable NLP measurement. This progress lends credence to the emerging evidence in the field of applied linguistics in the rising argument about multidimensional assessment procedures at the discourse level.

The pedagogical and technology implications are far-reaching. The DECOR may be applied under the digital platforms to provide the instant feedback throughout the writing process or act as an after-draft assessment tool. It is a modular system that can be customized to run through different levels of learners with different ability like detection, explanations, and guided rewriting. The study also confirms that DECOR has application to summative assessment conditions where a broader view of writing proficiency is more and more required. Such implemented examples indicate that DECOR may become a crucial element in personalized training and broad-scale education.

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

The study however has some limitations. The rewritings that were done by DECOR, although most of them worked, tended to be PC or very awkward in their expression or simplified the language in an awkward way especially when experiencing such language use as idiomatic or creative language use. Its explanation mechanism is to be improved as well to be pedagogically clear. There should also be an ethical concern about AI in education with considerations to bias, data security, and overdependence on automation. Further studies ought to be conducted on cross-linguistic uses, pattern of engagement by learners and long-term impacts on writing skills. Generally, the paper serves to fill the gap between computation linguistics and practice pedagogy, which makes the DECOR a useful tool in such an evolving environment of language learning.

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. *New Research on Cohesion and Coherence in Linguistics*, 182-201. https://doi.org/10.4324/9781003190110-10-13

Granger, S. (2021). 1 phraseology, corpora and L2 research. *Perspectives on the L2 Phrasicon*, 3-22. https://doi.org/10.21832/9781788924863-002

Hanaoka, O., & Izumi, S. (2021). Directions for future research on attention and L2 writing. *The Routledge Handbook of Second Language Acquisition and Writing*, 312-324. https://doi.org/10.4324/9780429199691-32

Hartwell, K., & Aull, L. (2021). Automated text-matching and writing-assistance tools. *Assessing Writing*, *50*, 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. *The Open Handbook of Linguistic Data Management*, 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. *2025 International Conference on Multi-Agent Systems for Collaborative Intelligence (ICMSCI)*, 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. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, 11436-11458. https://doi.org/10.18653/v1/2024.emnlp-main.639