A COMPARATIVE STUDY OF COHESION AND COHERENCE IN MALE AND FEMALE AUTHORED NOVELS JOYCE'S *A PORTRAIT OF THE ARTIST AS A YOUNG MAN AND* WOOLF'S *TO THE LIGHTHOUSE*: A CORPUS-BASED ANALYSIS

Sidra Yasin, Faizullah, Hira Haroon
1. M.Phil. Scholar, Department of English Graduate Studies, National University of Modern Languages
2. M.Phil. Scholar, Department of English Graduate Studies, National University of Modern Languages; Email: faizp057@gmail.com
3. Lecturer, Department of Applied Linguistics, GC University Faisalabad.

Abstract
This corpus-based study investigates the role of cohesion and coherence in literary texts, focusing on James Joyce's A Portrait of the Artist as a Young Man *and Virginia Woolf's* To the Lighthouse. *Drawing on Halliday and Hasan's (1976) cohesion theory, this research aims to encourage new scholars to explore corpus-based methodologies within literature. By applying Part-of-Speech (POS) tagging and collocation analysis, the study seeks to identify key linguistic patterns that contribute to the cohesion and coherence of the texts. The thirty most frequently used words from each text were analyzed, focusing on nouns, verbs, and adjectives, to highlight structural similarities and differences and reveal co-occurring linguistic elements. Using AntConc 3.4.4 software, these findings offer a quantitative lens on how language structures support textual coherence and thematic development within literary works. Employing a corpus-based quantitative research design, the study involved digitizing both texts for analysis. POS tagging and collocation analysis facilitated a systematic examination of linguistic patterns. The digital tool AntConc 3.4.4 enabled accurate identification and categorization of words by parts of speech, supporting comparative analysis between the two works. The study reveals significant patterns in noun, verb, and adjective use across the two texts, demonstrating that cohesion and coherence in literature can be effectively examined through corpus-based approaches. The analysis confirms that certain high-frequency words contribute prominently to the texts' thematic and narrative structure, supporting Halliday and Hasan's cohesion model in literary contexts. This research establishes that corpus-based techniques can provide valuable insights into literary texts, highlighting how cohesion and coherence manifest through specific linguistic structures. The study underscores the potential of corpus linguistics in literary analysis, encouraging future research to integrate these methods for a more comprehensive understanding of literary texts.*

Keywords: Corpus linguistics, cohesion and coherence, Halliday and Hasan (1976), part-of-speech tagging

1. INTRODUCTION

Corpus linguistics provides a structured way to analyze language by examining large pools of text, often written by a specific author or on a precise topic. Although the methods associated with corpus linguistics were conceptualized as early as the 1960s, the term itself only emerged in the 1980s. This approach has since proven valuable for both quantitative and qualitative studies, capturing nuances in language and data. Corpus linguistics is widely applicable across language studies, offering insights into language variations; however, its use in literary text analysis remains relatively underexplored. While corpus-driven approaches yield insightful data patterns, personal interpretation and knowledge of the text by the researcher are essential for nuanced analysis.

In this study, AntConc 3.4.4, developed by Laurence Anthony, was used to perform quantitative analysis on the texts, focusing specifically on nouns, verbs, adjectives, and common phrases. Analyzing these parts of speech—particularly the most frequently occurring words—helps uncover underlying patterns of cohesion and coherence within the texts. Notably, these

parts of speech (e.g., nouns, verbs, and adjectives) play critical roles in reading, writing, listening, and speaking by clarifying ideas and expressing emotions.

This study explores two renowned works by James Joyce and Virginia Woolf—A Portrait of the Artist as a Young Man and To the Lighthouse. Both novels offer a rich examination of inner worlds and external realities. Joyce's novel reflects on the protagonist Stephen Dedalus's journey toward becoming a writer, filled with inner growth and philosophical questioning, mirroring Joyce's own creative struggles. Woolf's To the Lighthouse follows the Ramsay family's experiences in Scotland, revealing each character's inner thoughts. Mrs. Ramsay represents kindness and love, contrasting with Mr. Ramsay's realism. Lily Briscoe, a young artist in the novel, symbolizes Woolf herself, grappling with creative expression and loss.

This research seeks to determine how cohesion and coherence manifest in these texts through Part-of-Speech (POS) Tagging and Collocation analysis, highlighting the interplay between language patterns and literary expression. The study not only advances corpus-based literary analysis but also underscores how linguistic tools can reveal deeper thematic connections within classic literature.

1.1. Statement of the Problem

Despite the growing application of corpus linguistics in language studies, its use in analyzing literary texts remains limited, particularly when exploring cohesion and coherence within classic literature. Traditional literary analysis often relies on subjective interpretation, which can overlook objective language patterns that reveal deeper textual cohesion. This study addresses this gap by applying corpus-based techniques, specifically POS tagging and collocation analysis, to examine how cohesion and coherence are established in James Joyce's A Portrait of the Artist as a Young Man and Virginia Woolf's To the Lighthouse. By focusing on these objective linguistic markers, the study aims to provide a comprehensive understanding of how linguistic elements contribute to the thematic and structural integrity of these texts.

1.2. Significance of the Study

This research holds significance for both literary scholars and corpus linguists. By integrating corpus linguistics with literary analysis, it offers a novel methodological approach for examining classic texts, contributing to the growing interdisciplinary field of digital humanities. For literary scholars, this study provides insights into how authors like Joyce and Woolf use language to create coherence and cohesion, enriching our understanding of their narrative techniques and stylistic choices. For corpus linguists, it expands the application of corpus-based methods to literature, demonstrating how these tools can be used beyond standard linguistic research. Ultimately, this study not only enhances our comprehension of these two iconic works but also encourages further corpus-based exploration of cohesion and coherence in literary texts, paving the way for new research in both fields.

1.3. Hypothesis:
The hypothesis of this study is that male authors exhibit greater cohesion in their writing compared to female authors.

1.4. Research Questions

This study aims to address the following questions through corpus-based analysis:

1. To what extent does Part of Speech (POS) Tagging capture coherence and cohesion in *A Portrait of the Artist as a Young Man* and *To the Lighthouse*?

2. How effectively does collocation analysis reveal coherence and cohesion within these two novels?

1.5. Research Objectives

The objectives of this research are as follows:
a. To highlight the coherence and cohesive relationships within both novels.
 b. To examine the impact of cohesion and coherence on the texts through collocation analysis.
c. To explore the grammatical and lexical characteristics of the novels using POS Tagging.

2. LITERATURE REVIEW

Since the last few decades we have seen that the corpus-based approaches in applied linguistics highlight the use of linguistic strategies in literary and non-literary texts. It assists the scholars to go forward with the correct statistics and co-occurrence of linguistic foundations and this thing bestows individuality to any writer's work. Hence, it is thought to be a scientific approach because the results are more trustworthy and quantifiable. Selected corpus linguistic experts like Mahlberg (2005) and Bonelli say that the kind of corpus-based research lets the researchers examine their assumption on confident linguistic and documented structures and the domino effect acquired in the usage of computable data. And this can be described semantically in reckoning out the text style of the text. These performances either reckonable or qualitative can outcome in form of detailed description of any literary text.

Noguchi (2004) emphasizes on the significance of AntConc software 3.4.4 and he further highlights that it is an informal way to custom this software for corpus scrutiny. It is insubstantial, modest and stress-free to use. It can also be used easily by the students anywhere like in the classroom and at home. McEnery is of the view that corpus linguistics can be seen as a method of how we can use corpora while teaching language, learning or in language studies also.

Kuebler and Zinsmeister (2015) say in their notable work "Corpus Linguistically Annotated Corpora" that corpus linguistics can be a tool or a theory at the same time. It is its use that decides how to apply it. Faiz Ullah and Dr. Maimoona Abdulaziz (2023) in their study, "The Use of Cohesion in William Golding's Lord of the Flies," apply a corpus linguistics approach to examine how grammatical cohesion, specifically through the use of conjunctions, contributes to the narrative coherence of the novel. Utilizing Halliday and Hassan's (1976) model of cohesion and the AntConc3.4.4 corpus tool, the study analyzes the frequency and distribution of conjunctions like 'and,' 'but,' 'so,' and 'yet' to understand how these markers link ideas and structure the text. The research underscores the importance of conjunctions in shaping the flow of the narrative and enhancing its overall meaning, offering a more transparent and objective analysis of the linguistic features that bind the text together.

Semino and Short (2004) are of the mark that corpus-based analysis is no doubt purely quantitative but it does not mean that the qualitative analysis cannot be made or should be excluded. They say if both are diversified they can comfort us to get a sophisticated level of consideration of literary work and selected pieces of literature. Um-e-Ammara, Rehana Yasim Anjum & Maryiam Javed (2019) in their article 'A Corpus-Based Halliday's Transitivity Analysis of *To the Lighthouse* (University of Management and Technology, Sialkot)' have explored transitivity in the novel To *the Lighthouse* by Virginia Woolf. They have done clause analysis of the novel. They did this transitivity analysis by using Antconc 3.5.7 corpus analysis tool software. They first tagged the text and then analyzed. This was totally a clause analysis in which they bring forth all types of processes and circumstances. They did their study under the domain of Corpus linguistic which is computer centered for the textual and database analysis of any text.
Mingzhu Zhao (2012) in her article The Art of Balance: 'A Corpus-assisted Stylistic Analysis of

Woolfian Parallelism in *To the Lighthouse* (Minzu University of China)' has examined thickness and parallelisms in Virginia Woolf's novel *To the Lighthouse*. Her study is sample-based judgment between this novel and other novels of the time to find out the specific philological and syntactic configurations that is the specification of wolf's parallelism. She has done her analysis through corpus-assisted construing and literal exploration of her work. She demonstrates Woolfian parallelism is defined by etymological bundles, contrasted prepositional phrases, -ing participles and oppositional arrangements.

NITA JAIN (2010) in her article 'Stylistic Techniques in *To The Lighthouse*' tells about the stylistic techniques of Virginia woolf. Nita is of the view that Woolf uses stream of consciousness, omniscient narrative, highly punctuated sentences, metaphors repetition, allusions, similes, imagery, complex structure. According to Nita Woolf uses third person person omniscient technique as commentator does in the match and this technique gives full access into thoughts, feelings, musings, and motivations of all characters.

Talking about James Joyce's novel *A Portrait of the Artist as a Youngman* Muhammad Ajmal and Aisha Shoukat in their article 'A Corpus-Based Study of James Joyce's *A Portrait of the Artist as Youngman* (University of Lahore) have utilized corpus-based technique to make the analysis of the novel. Their main aim of this research is to promote this style. They both have tried their best in this article to tell the researchers how this technique can be helpful and good result making for them in upcoming time. They collected thirty words from the novel and their frequency was noted. They used AntConc software to interpret and explain the text.

Shmoop Editorial (2008) Team in its article *A Portrait of the Artist as a Youngman writing Style* (Shmoop University 2008) tells about free indirect discourse that this technique how the narrative voice takes change between the character's minds and the outer world of the novel. Moreover, free indirect discourse is narrative style that mixes first person narrative with a third person narrative in the character's mind. It is this style that urges Joyce to keep some distance from his character while Joyce is revealing his innermost working to us. This thing also allows the author to practice little irony to clear things. But it does not mean that Joyce is humor writer.

Jonathan Reeve (2015) is his article A Macro-Etymological Analysis of James Joyce's *A Portrait of the Artist as a Youngman* brings to light that the English language has been influenced by many languages it came into contact with like French, Latin, Ancient Greek and Irish. He says when a writer chooses a word 'chew over masticate' and 'enchantment over spell' this choice indicates something. To measure these stylistic vectors, he uses calculus breakdown of the derivations of the words in James Joyce's novel *'A Portrait of the Artist as a Youngman'* to pinpoint etymological register, resonances, and the level of discourse. He also has brought to light how these original signals reflect physical elements of the novel.

Above are some different styles and perspectives in which many researchers have analyzed both the novels. But I intend to dissect these novels in the perspective of Corpus-based analysis to identify cohesion and coherence through Collocation and POS Tagging.

3. RESEARCH METHODOLOGY

In this research Mahlberg and Mclntyre's (2012) methodology has been used that brings to light that a piece of text can also be the part of corpus. This research is both qualitative and quantitative. The major source of figures collection is the internet. After scrutinizing different articles and software to do analysis of the novels, *A Portrait of the Artist as a Young Man* and *To the LightHouse* AntConc software was used. To bring to light the coherence and cohesion

between the both texts Part of speech (POS) Tagging and Collocation in the texts have been observed. The numbers of nouns, verbs, adjectives, words and phrases have been analyzed. These features give linguistic coherence and cohesion to the texts. Thirty most frequently used words have been selected as a part of speech to attain the results. The novels were analyzed in very short and limited time. There is no doubt about it that research's personal experience and knowledge supports a lot of analysis but this research also brings forth easier ways to collect data and analyze it by laptop or PC than manually.

For the purpose of research two novels are collected from the internet one is written by male author A portrait of the artist as a young man by Joyce and another one is written by female author To the lighthouse written by virginia woolf.
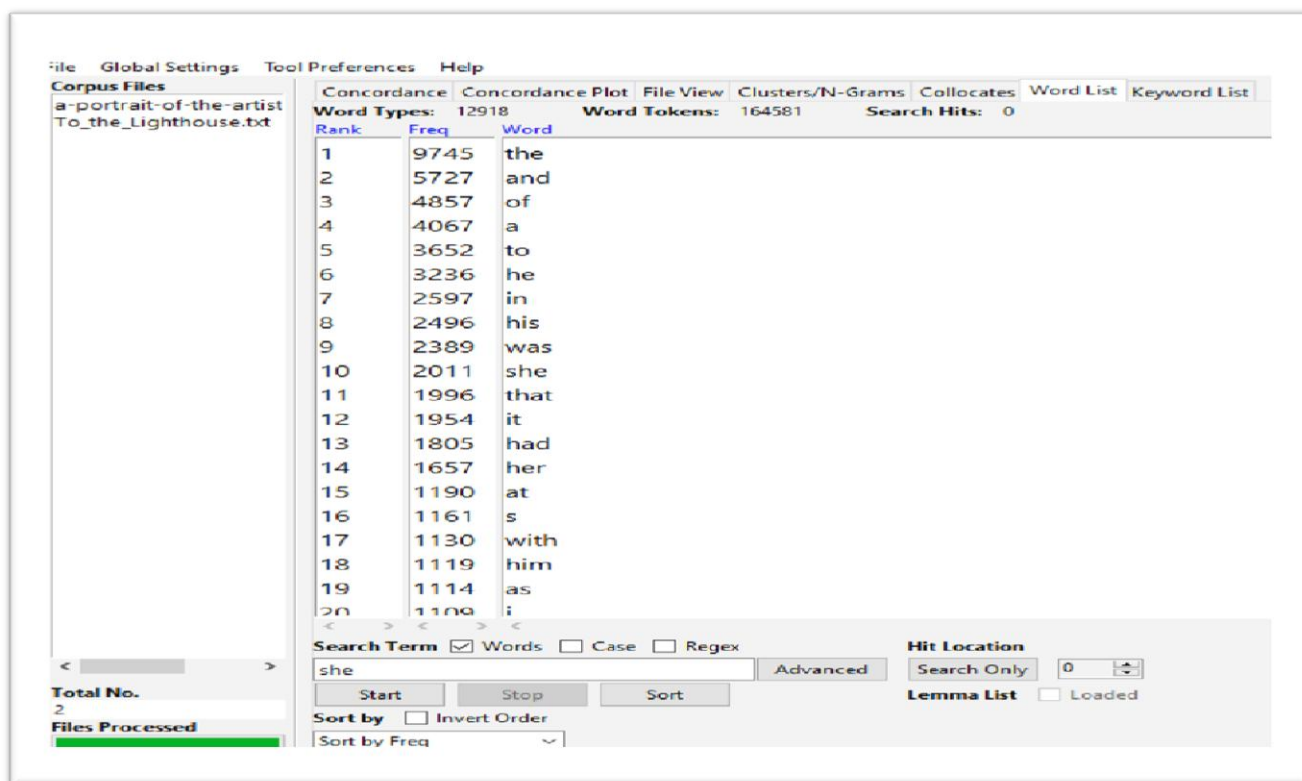
4. DATA ANALYSIS

Figure 1: Wordlist of 20 frequently used words

Figure 1 shows words that are repeatedly recycled in both the novels. The words are not pertaining to one part of the speech. The words belong to different parts of speech like adjectives, verbs, nouns, and articles. The frequency displays how many times the word has been used in the both novels. The article' the' and 'a' is used 9745, 4067 times in the novels. The verb was, had used 2389, 1805 times in both novels. The pronoun he, his used 3236, 2496 times. The next figure no 2 shows 10 most frequently used words in the novel.

Figure 2: Wordlist of 10 frequently used words

Rank	Freq	Word
20	1109	i
21	1098	for
22	1022	on
23	1020	said
24	960	they
25	946	not
26	853	you
27	815	but
28	791	t
29	761	all
30	741	he

Word Types: 12918   Word Tokens: 164581   Search Hits: 0

The frequency of the word the figure are the pronouns I, they, you is used 1109, 960, 853 times. The preposition for, on is used 1098, 1022 times. The conjunction is used 815 times in novels. The verb be is used 741 times in novels.

A discussion on the most frequent word list in the context of cohesion and coherence analysis reveals critical insights into narrative style and structure. In the comparative study of *A Portrait of the Artist as a Young Man* by James Joyce and *To the Lighthouse* by Virginia Woolf, frequent words such as "he," "was," and "the" were highlighted, reflecting different parts of speech, including nouns, verbs, pronouns, and articles. The frequency and role of these words demonstrate each author's stylistic tendencies and thematic focus. Joyce's usage shows a clear, cohesive structure centered around particular verbs and articles, aligning with his focus on the protagonist's inner development. Conversely, Woolf's frequent use of connectors like pronouns and prepositions builds thematic coherence, depicting complex relational dynamics and inner reflections of her characters. This frequency analysis, facilitated by AntConc software, showcases how quantitative data can illuminate stylistic elements, providing a detailed view of how each author achieves coherence and cohesion in their narratives. This approach offers a model for future corpus-based literary analysis, emphasizing the interpretive power of word frequency data to uncover underlying narrative patterns.

Male writing	Frequency	Female writting	Frequency
Noun (he)	1878	Noun (he)	1358
Article (the)	6070	Article (the)	3675

Verb (was)	1066	Verb (was)	1323
Preposition (on)	538	Preposition (on)	484
Conjunction (but)	356	Conjunction (but)	459
Pronoun (his)	1745	Pronoun (his)	751
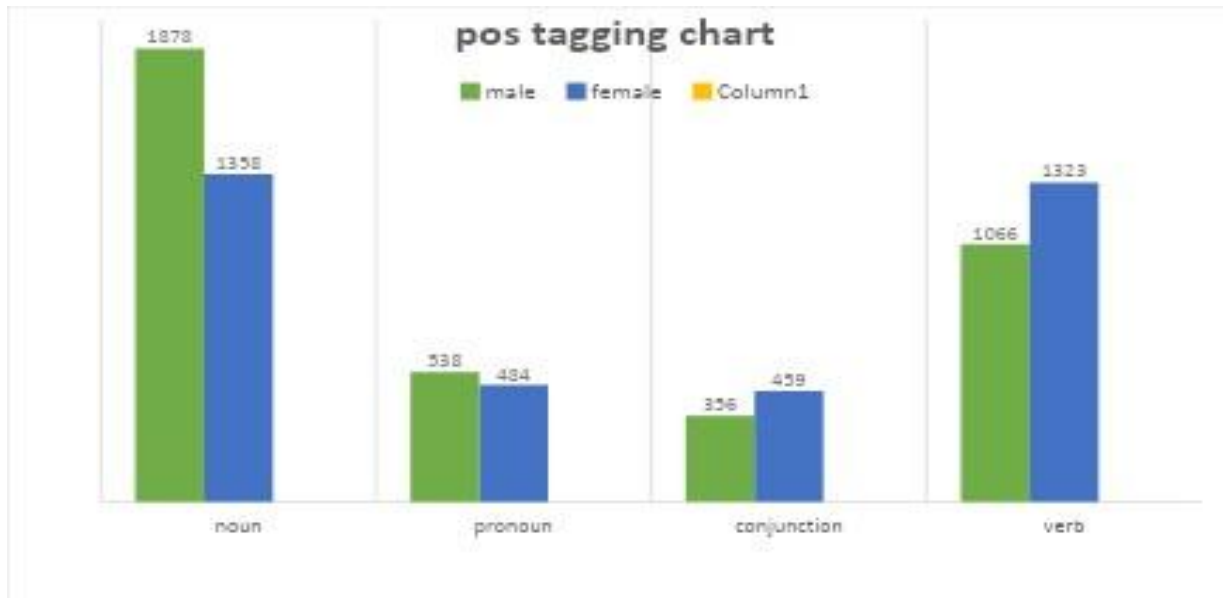
5. RESULTS AND DISCUSSION

AntConc software was used to assess coherence and cohesion in *A Portrait of the Artist as a Young Man* by James Joyce and *To the Lighthouse* by Virginia Woolf. The table captures various parts of speech—such as nouns, verbs, pronouns, and articles—highlighting their frequency and patterns in both texts. For example, nouns like "he" and pronouns like "his" are used differently across the novels, reflecting distinct stylistic and narrative techniques by each author. Male author Joyce shows a greater emphasis on certain verbs and articles, contributing to a cohesive structure that aligns with his narrative style. In contrast, Woolf's usage presents more variation in connectors, suggesting a style rich in thematic interconnections. This comparison indicates how male and female authors employ coherence and cohesion differently, potentially influenced by narrative goals and thematic focus. Overall, the table underlines the significance of quantitative analysis through AntConc in understanding linguistic features and stylistic choices that contribute to the distinctive coherence in each literary work.

6. CONCLUSION

This study indicated Corpus Linguistics as a strong apparatus in the inquiry of literary texts, particularly when up mixed with no computational tactics to highlight the details of important words in a text. Corpus linguistics assists researchers and provided them perfect approach to study various qualities of texts that remained unseen. It enables the researchers that how to manage and process the large quantity of data to give statistical treatment to texts. This study also emphasized the importance of using AntConc Software for the corpus-based analysis of the literary texts, *A Portrait of the Artist as a Youngman* by James Joyce and *To the lightHouse*, by Virginia Woolf. To find out the coherence and cohesion in the both text the noun, verb, adjectives (POS Tagging) and word and phrases (Collocation) have been analyzed. Hence the most habitually used words in the both novels were inspected. The fallouts specify that these words belong to diverse part of speech. This exploration paves the way for upcoming researchers who want to discover new dimensions in this sphere of literature. Through AntConc software 3.4.4 toolkit the researchers can classify frequency, wordlist, concordance, POS Tagging and Collocation.

REFERENCES

1. Ammara, U, Anjum, R, Y & Javed, M. (2019) '*A Corpus-based Halliday's Transitivity Analysis o*f *to the Light House* (University of Management and Technology, Sialkot)
2. Hunstons, S. (2002). *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
3. Zhao,M.   (2012). *The Art of Balance*: *A Corpus-assisted Stylistic Analysis* of Woolfian Parallelism in *To the LightHouse* (Minzu University of China).
4. Kuebler, Zinsmeister. (2015). *Corpus Linguistic and Linguistically Annotated Corpora*.
5. Jain, N. (2010) '*Stylistic Techniques in To the LightHouse'*
6. Mahlberg, Michaela. (2005). *English general nouns: A corpus theoretical approach.*
7. McEnery, Tony, Richard Z. Xiao, and Yukio Tono. (2005). *Corpus-based language studies: An advanced resource book.* London: Routledge.
8. Shmoop Editorial (2008). Team *A Portrait of the Artist as a Youngman writing Style* (Shmoop University)
9. Mcintyre, D. (2015). *Towards an Integrated Corpus Stylistics.''* Topic in Linguistics:59-68.
10. Reeve, J. (2015) *A Macro-Etymological Analysis of* James Joyce's *A Portrait of the Artist as a Youngman*
11. Semino, Elena, and & Mick Short. (2004). Corpus Stylistics: Speech, Writing, and Thought presentation in a corpus of English writing. London: Routledge, Print
12. Noguchi, J. (2004). A *genre analysis and mini-corpora approach to support professional writing by non-native speakers. English Corpus Studies,* 11,101-110
13. Javier Lopez. (2021). *Discourse analysis cohesion and coherence: studying cohesion and coherence in discourse*. Sierra university. journal of artificial intelligence, vol (1)
14. ULLAH, F., & Abdulaziz, M. (2023). A CORPUS-BASED ANALYSIS OF COHESION IN WILLIAM GOLDING'S FICTIONAL REALMS: LORD OF THE FLIES. *Jahan-e-Tahqeeq*, *6*(3), 440-464.
15. Wang Gang, & Liu. (2014).On the theoretical framework of the study of discourse cohesion and coherence.Studies in literature and language.vol(8). DOI:10.3968/4512