Pakistan Journal of Language Studies
Vol. 9, Issue 1, 2025

ISSN (Online) 2664-1461
ISSN (Print) 2664-1453

Lexical Cohesion and Gender in Intermediate-Level Student Writing: An NLP-Based Statistical Case Study

Nida Ashiq[1], Dr. Muhammad Ahmad[2]

[1]MPhil Scholar, Department of English, University of Okara, Okara, Pakistan.
Email: nidach1696635@gmail.com

[2]Visiting Lecturer, Department of English, University of Okara, Okara, Pakistan.
Corresponding Author: ahmad453@yandex.com

Abstract

Cohesion is a crucial component of academic writing because it helps to make a text coherent and logical. The purpose of the present study is to examine the lexical cohesive devices used by intermediate level students and to shed light on gender difference in their use in student's written compositions. The research is based on Halliday and Hasan (1976) theory of cohesion which is developed in lexical branch such as repetition, synonymy, antonymy, superordination, subordination and collocation. Both descriptive and inferential statistics were used for data analysis, and a quantitative research design was used.The Corpus consisted of 30 paragraphs produced by the students of FSc 1st Year (Morning) of Punjab College, Okara. An automated Natural Language Processing (NLP) pipeline using Python was created for annotation, which includes tokenisation, POS Tagging, Lemmatisation, WordNet semantic matching, for accurately identifying the six lexical categories. The results showed that there were a lot of differences between the use of the lexical cohesive devices of males and females. Female students used more cohesive ties than male students across all six categories (479 vs. 234), with statistically significant differences for repetition ($t(28) = 5.16$, $p < .001$), synonymy ($t(28) = 2.15$, $p = .040$), and collocation ($t(28) = 2.92$, $p = .007$). The most common categories with males were repetition and collocation (57.7%), whereas with females they were repetition (67.2%). Subordination and superordination were the least common categories of the devices used. A one-way ANOVA showed that there was significant difference between categories, $F(5, 174) = 27.83$, $p < .001$, for which collocation was found to be a statistically different category from reiteration. These results were further validated by the experts, who found that the writing scores of the female students (7.97) were higher than those of male students (7.12).The findings indicate that women students are more proficient in the domain of lexical cohesion which in turn helps to increase the text coherence. This study has highlighted the need of explicit teaching of cohesive devices and provided some implications for the language teachers, the curriculum design and future researchers.

Keywords: lexical cohesion, POS tagging, NLP annotation, WordNet, gender differences, intermediate-level writing

1. Introduction

Writing occupies a central place in second- and foreign-language education. The reason is that a learner can express his/her ideas in the form of writing by producing

articles, different types of texts, and essays as their projects and assignments etc. As Paltridge and Starfield (2007) have narrated this view that academic writing is measured to express the ideas of students in the degree program as well as it is very helpful to understand the different patterns of writing. When students are affiliated to an institution then they are bound to fulfill the demands of 'writing education' because academic writings provide a real glimpse of a language. In academic writing, students present their views on paper by following the basic rules of writing. The rules of writing can be in the form of cohesive elements along with the elements of coherence which is the demand of present study. The use of cohesive devices to join different words, phrases, and sentences in the formation of text is important. The study of how texts are put together, commonly termed text linguistics, treats the organisation and production of a text as an object of analysis in its own right, rather than as a vehicle for content alone

Academic writing in particular requires learners to master the conventions that hold a text together: ideas need to be logically sequenced, related information grouped, unnecessary repetition avoided, and the intended meaning kept unambiguous (Paltridge & Starfield, 2007). Two properties are repeatedly identified as central to meeting these conventions: cohesion, the linguistic linking of sentences and paragraphs, and coherence, the underlying logical unity of the ideas being expressed.

The two properties operate at different levels of the text. Cohesion works at the surface, or micro, level, using devices such as reference, substitution, ellipsis, conjunction, and lexical ties to bind sentences together (Halliday & Hasan, 1976); coherence works at the macro level, concerning whether the ideas expressed hang together conceptually regardless of the surface ties used to express them (Crystal, 2006). A paragraph that strings together unrelated statements, for example, one sentence describing a teacher writing on a board, a second describing a car stopping near a market, and a third describing children flying kites in a park, fails on both counts there is no shared referent linking the sentences and no single idea that the paragraph as a whole develops. Cohesive ties supply the missing linkage; Halliday and Hasan (1976) illustrate this with a simple example in which a pronoun such as it or him cannot be interpreted without reference to an earlier sentence, and it is precisely this dependency that gives a text what they term its texture and that distinguishes a text from a mere sequence of sentences (Brown & Yule, 1983; Connor, 1984).

The ability to link ideas smoothly across sentences is one of the clearest markers of writing proficiency, and its absence is one of the most common complaints teachers raise about student prose (Witte & Faigley, 1981). This capacity rests on the network of ties described above, and subsequent corpus-based work has repeatedly linked their density and variety to independent judgements of text quality (McNamara, Louwerse, McCarthy, & Graesser, 2010). Of the two branches of the model, lexical cohesion, created through repetition, semantically related vocabulary, and habitually co-occurring words, has proved particularly productive for classroom-based research because it is directly observable in a learner's word choices rather than inferred from more abstract grammatical relations.

1.1 Statement of the Problem

It is generally observed that students have to face great difficulty in exams due to lack of cohesion in the text which affects their grades. Most of the students do not know the usage of these devices accurately, like they do not have adequate understanding of grammatical structures and insufficient knowledge about cohesive devices. This lack of knowledge about cohesive devices creates a significant problem to produce a coherent text at all, resulting in disjointed and fragmented writing which can break good academic performance. Another problem in this realm is to understand the role of gender which has an impact on writing skill related to its quality and style. More specifically, teachers at the intermediate level in Pakistani colleges routinely observe that student paragraphs can be grammatically accurate while still reading as disconnected, and that male and female students appear to differ in how well their ideas hang together. This observation, however, has typically been reported impressionistically rather than tested: prior accounts of lexical cohesion in learner writing have relied on manual, colour-coded annotation carried out by a single researcher, a procedure that is time-consuming, difficult to replicate exactly, and rarely accompanied by inferential statistics capable of distinguishing a genuine gender difference from ordinary sample variation. The problem this study addresses is therefore twofold: first, there is no replicable, automated account of which specific categories of lexical cohesion repetition, synonymy, antonymy, subordination, superordination, and collocation male and female intermediate-level students actually draw on; second, where gender differences in these categories have been reported, they have rarely been tested for statistical significance, leaving open the possibility that some or all of the reported differences would not survive formal testing.

1.2 Research Questions

This study is organized around one central research question, with two statistically operationalized sub-questions that follow from it:

RQ1: What categories of cohesive devices are employed by male and female intermediate-level students in their written compositions?

RQ1a: Do male and female students differ significantly in their frequency of use of each lexical category, as tested by independent-samples t-tests?

RQ1b: Does frequency of use differ significantly across the six lexical categories themselves, as tested by one-way ANOVA, independent of the gender comparison?

1.3 Significance of the Study

The importance of the study lies in enhancing educational insights as it will be very helpful for teachers to adopt effective pedagogical strategies to improve writing skills focusing on gender. This study will play a vital role in improving the academic performance of students in the realm of writing and even in communication. This study will perform a significant role on the issue of gender equality as it will contribute to creating an inclusive environment related to writing skills between genders. This research

study contributes to enhance the knowledge about the study of language and gender. The study is also significant on three additional levels. Theoretically, it tests Halliday and Hasan's (1976) category-level model of lexical cohesion against a Pakistani intermediate-level corpus using an automated, rule-based procedure, adding to a literature that has so far relied almost entirely on manual annotation. Methodologically, it demonstrates that a POS-tagging-based NLP pipeline, combined with formal inferential statistics computed in Python, can reproduce and refine findings that earlier studies established only through slower, harder-to-replicate manual coding, offering other researchers a transparent and reusable procedure rather than a one-off hand annotation. Pedagogically, the study is significant for teachers and curriculum designers at the intermediate level: by identifying precisely which categories show a statistically reliable gender gap (repetition, synonymy, and collocation, in this class) and which do not (antonymy, subordination, and superordination), it points instruction toward specific, evidence-based targets rather than a general appeal to "improve cohesion," and toward categories that both sexes underuse and could usefully be taught to the whole class rather than to one gender alone.

1.4 Objectives of the Study

- It aims to identify the types of lexical cohesive devices in written text of male and female students.
- The study aims to determine the usage of lexical cohesive devices in written text of students on the basis of gender.
- The other aim of this research study is to provide a comprehensive understanding about the role of gender in the composition of text by using lexical cohesive devices.

2. Review of Related Literature

Halliday and Hasan (1976) divide cohesion into grammatical ties (reference, substitution, ellipsis, conjunction) and lexical ties, the latter split into reiteration (repetition, synonymy, antonymy, hyponymy, and superordination) and collocation. This lexical branch has been applied productively to learner corpora because lexical choices are more easily identified and quantified than grammatical relations, and because they correlate strongly with perceived text quality (Mahlberg, 2006).

Cohesion is defined as the set of linguistic resources that link sentences and paragraphs together to create a unified text. Halliday and Hasan (1976) argue that a text has "texture" precisely because of these cohesive relationships, and it is this texture that distinguishes a text from a random collection of sentences. Cohesion operates through the semantic relationships between elements in a text, where one element is interpreted as a function of another.

Text and cohesion are closely linked to each other. Trebits (2009) narrates that cohesion contributes to generate a logical and integrated text with the help of words and phrases and their relationship with lexical items. Significance of cohesion is emphasized by this author in a vivid and joined written narrative. According to Gutwinski(1976)

cohesion is a relationship between different phrases and sentences that are responsible for the composition of text. The relationship between these elements has two layers: the first one is grammatical layer and the other is lexical layer that have the reflection about the structure of semiotic discourse.

Tools of cohesion that are used for the comprehension of a text can be illustrated by defining the relationship between different elements on semantic levels. Two elements can have a connection between each other on semantic level can exist in the text and one of both can be existed out of the text (Witte & Faigley, 1981). If two elements existed in the text, it comes under the umbrella of endophoric reference, and if one element comes in the text and other is outside of the text, this situation comes under the umbrella of exophoric reference. These two types of references are the part of cohesion itself.

Lexical cohesion is concerned with the semantic relationships between words and phrases within a text. It is established through the careful choice of vocabulary that links sentences and ideas together, ensuring that the text is unified and meaningful. According to Matthiessen (2014), the selection of vocabulary is a crucial tool for creating meaning and conveying the intended message.

Reiteration is considered a significant means of lexical cohesion and is a very important element in the construction of text in academic writing (Adorjan, 2015; Halliday & Hasan, 1976). Reiteration is not only the repetition of the same words, but also the repetition of phrases and ideas in related forms to reinforce the writer's position, to clarify the text, and to lead the reader through the text. Tanskanen (2006) noted that repetition and reiteration differ: the latter is broader than the former, as it can include synonyms, hyponyms, superordinates, or general words. Studies have consistently shown that reiteration rescues a text from monotony, increases comprehension, and enhances readability (Idegbekwe, 2022).

Repetition can be defined as the recurrence of the same word or expression to maintain the reader's attention on a particular idea. It is more commonly used in narrative essays, as writers employ repetition to emphasize the central focus of their topic. For example: Being an only child offers a life like that of a princess. I have always found it amusing when people assume that being an only child must be lonely and boring.

Synonyms are words that have the same meaning and can be used together in a sentence without significantly altering the overall meaning (Purba et al., 2023). Using synonyms helps reduce unnecessary repetition and attracts the reader's attention (Sumani & Kartikasari, 2022). Near synonyms, also known as plesionymy, refer to pairs of words that are similar in some secondary aspects but differ in their primary or fundamental meanings (Hameed & Challob, 2022).

Antonyms are an important lexical cohesion device that involves the use of words or expressions with opposite meanings. They are commonly used in discourse to create contrast, emphasize ideas, and highlight differences between concepts. According to Halliday and Matthiessen (2014), antonyms help establish textual connections by creating a sense of balance or tension within the text.

Subordination (Hyponymy) refers to the use of a more specific term to refer back to a more general one already introduced. Superordination (Hypernymy) is the reverse relation a more general term referring back to a specific one. Paltridge (2012) describes superordinate as a type of repetition in which a more general term is used to refer to a previously defined specific object. For instance, the word “fruit” in a text can refer to more specific terms like “apple,” “banana,” and “mango” (Sari & Anindita, 2020).

Collocation is the association of words that are commonly found together and have a natural link in language. Unlike other kinds of cohesive devices, collocations are not arbitrary; they are culturally and linguistically "natural" and "acceptable" to native speakers. For example, the words "ocean," "fish," and "beach" are collocated because they belong to the same semantic field related to the marine environment.

There are different types of cohesion in which lexical cohesion is used in written form along with references and conjunction; on the other hand substitution and ellipsis are used in oral form (De Beaugrande & Dressier, 1981: Witte & Faigley, 1981). According to Aytas (2008), coherence is another element that focuses on text in text linguistics. It comes after the cohesion and it is essential for the composition of text. The text can be taken as incomplete and insufficient. A good text can be produced if it has tools of cohesion along with semantic coherence. If a text is having comprehensive details, succeeding and proceeding statements, and there is no contradicting elements in the text, then it can be called a text having 'coherence'. Cohesion is the connection between different sentences on grammatical level for the formation of text, but on the other hand coherence is the connection between different sentences on semantic level (Aksan, 1999). Cohesion explains the mechanisms that are responsible to bring the text together, but coherence explains those mechanisms on semantic level which are used to make text understandable (Witte & Faigley, 1981).

The written text of students can be evaluated on the basis of formal property like: grammar, writing and punctuation, but it should be evaluated by keeping an eye on different property like logical coherence; adequate changing between paragraphs, basic idea along with assisting ideas, and the last part of the text must be inclusive and impressive. There are so many students have to face too many difficulties to express their ideas on the paper, which are existed in their mind, even after completing their secondary education (Can, 2012: Karatay, 2010). So many problems have to face while producing and arranging different ideas, present in the mind and expressing them in written form (Richards & Renandya, 2002).

Halliday and Hasan (1976) stated that cohesion and coherence are the two essential elements to produce good writing. There are different research studies in which different researchers have given a lot of attention to cohesion while producing a text. Different learners have to face many problems in producing a coherent text and a text can be called a coherent text if there is 'Cohesion' in the text. So, 'Cohesion' is considered as a crucial sub-field of ESL writing. There are many research studies (Jafarpur, 1991, John, 1980, Johnson, 1992; Zhang, 2000; Hartnett, 1989) about Cohesion and Coherence, applying the framework of Halliday and Hasan (1976) has been done in second-language writing like ESL/EFL.

A consistent finding across this literature is that repetition dominates learner writing, with more demanding categories such as hyponymy and superordination trailing far behind. However; in their study Khalil & Abu-Ayyash (2023)examined the use of lexical cohesive devices (LCDs) in the writing of 56 Arab EFL learners in the UAE. The results indicated that the device most used was repetition (57.07% of the cohesive ties), followed by the use of collocations, and while the least used devices were hyponyms and meronyms. A moderate positive correlation ($r = 0.330$) was also found between the use of LCD and writing scores. In addition, qualitative findings revealed some problems such as weak knowledge of the students on some of the cohesive ties, lack of instruction focus, lack of teacher feedback, and L1 interference, which resulted in overuse of repetition. Based on their analysis, the researchers recommended incorporation of lexical cohesion into the writing program as a whole and not as discrete vocabulary items.

Sihombing et al. (2023) studied lexical cohesion in the background part of 10 student theses produced by the students of the HKBP Nommensen University. This study examined five types of lexical cohesion, which are: repetition, synonymy, antonymy, hyponymy, and meronymy. A total of 333 instances were found, with repetition being the most used (91.6%) followed by synonymy (0.9%), antonymy (2.1%), and hyponymy (2.7%) and meronymy (3.0%). Findings indicate that learners use repetition extensively and, as a consequence, do not show much lexical variation and thus, it is not that exciting to read. This study advises that students use more synonyms and antonyms in their writing, departments offer training on cohesive writing, and further research be conducted on other genres and grammatical cohesion for a more thorough analysis.

Wahyono et al. (2020) reported a comparable ranking repetition first, followed by hyponyms, synonyms, hypernyms, and antonyms in journalistic writing.

A smaller strand of work has treated gender as an explicit variable. Argamon et al. (2003) found that women's formal writing carried more of certain linguistic features than men's; Aprianto et al. (2024) found that female students favoured content-based paraphrasing while male students favoured syntactic restructuring; and Wulan Anggraeni et al. (2025) found that girls' narrative writing was more densely organised and more interpersonally elaborated than boys'. Against this trend, (Vali & Kianiparsa (2010) found no significant gender difference in cohesive device use, and Kaçauni and Qafzezi (2025) found that formal instruction, rather than gender, was the stronger predictor of learners' command of cohesive devices. Almost none of this literature, however, reports inferential statistics t-tests or ANOVA alongside the descriptive counts, which leaves open whether the reported gender differences would survive formal significance testing. The present case study was designed specifically to close that gap for a single, well-defined class.

This repetition-dominant hierarchy has been replicated across a wide range of text types and learner populations. Sidabutar (2021) found lexical cohesion accounting for the majority of cohesive ties in students' narrative writing, with hyponymy the least used category, and attributed the pattern to limited vocabulary and reliance on first-language equivalents; Bobihoe et al. (2025) reported a similar reliance on reiteration in Indonesian EFL academic writing, describing the resulting prose as monotonous and tracing it to the same underlying vocabulary gap. Narrowing to individual sub-categories, Hameed and

Challob (2022) found that Iraqi university students used synonymy to sustain logical coherence but frequently selected an inaccurate near-synonym for want of a wider vocabulary, and Sinaga et al. (2022) found repetition dominant in journalistic writing with hyponymy and antonymy each occurring only once.

Work specifically on collocation has identified a further layer of difficulty: Nguyen (2024) found that EFL essay writers favoured adjective-noun collocations over adverb-related ones and frequently misselected a component word, and Do and Le (2023) traced comparable verb-noun and adjective-noun collocation errors among Vietnamese learners to first-language interference. Together, these studies suggest that the repetition-heavy, hyponym-sparse profile reported for this study's corpus is not a peculiarity of the Pakistani intermediate context but a pattern that recurs wherever learner vocabulary, rather than instruction, is left to drive lexical choice.

A second strand of research has asked whether a denser use of cohesive devices actually predicts better writing, and the answer has been decidedly mixed. Several studies report a positive association between cohesive density and quality ratings (Ferris, 1994; Liu & Braine, 2005; Field & Oi, 1992), while others find no reliable relationship at all (Jafarpur, 1991; Castro, 2004; Johnson, 1992; Zhang, 2000). Saudin (2019) offers one explanation for this inconsistency: across two EFL proficiency levels, repetition was the most overused sub-category of lexical cohesion, and its overuse was associated with lower, not higher, quality ratings compared with collocation and synonymy, echoing Irwin's (1982) earlier observation that novice writers often use cohesive devices heavily without using them coherently. This body of work motivates the category-level, rather than purely aggregate, analysis adopted here: a single overall cohesion count would obscure the possibility that a high frequency of one category (repetition) coexists with a genuine shortage of others (synonymy, collocation), which is precisely the pattern the results in Section 5 describe for this class.

However, different researchers have different findings, but some have similar findings and many have concluded that there is no much difference between the use of cohesive devices while producing good writings as well as weak writings (Johnson, 1992; Zhang, 2000). The high frequency of cohesive devices does not have strong effects on writing quality and it challenges the notion that effective writers use more devices than that of less proficient writers. Different research studies present that writing proficiency is based on appropriate use of cohesive devices not on excessive use of cohesive devices. Further investigation can be made about the relationship among writing quality, cohesive devices, and language proficiency.

Jafarpur (1991) has given the concept of high-rated essays and low-rated essays on the basis of cohesive devices, as if there is high use of cohesive devices then the essay is considered as a high-rated essay and if there is low use of cohesive devices then it is considered as a low-rated essay. Some research studies have proved that a composition comprising more cohesion is having high scores than that of low scored composition, based on less cohesion (Jafarpur, 1991). These research studies have presented a positive correlation between the quality of writing and cohesion. Cohesion is directly proportional to scores of a composition. These scores increase and decrease in the same direction as:

high level of cohesion leads towards high scores and low level of cohesion leads towards low scores. It is obvious that the disjointed and fragmented writing is not easy to comprehend and proves as a hinder in effective communication. The significance of this aspect leads towards writing pedagogy that teachers should focus on cohesive strategies to improve writing qualities and to enhance powerful communication. Cohesion has priority in the realm of writing so teachers should enable their students to compose high-scoring texts by guiding them about the appropriate use of cohesive devices.

The intermediate level, the focus of the present study, has also emerged as a particularly responsive site for cohesion instruction. Alfalagg (2020) found that structured teacher-student writing conferences produced a large, statistically significant improvement in the frequency and accuracy of cohesive device use among intermediate-level EFL students, with referential markers and conjunctions showing the clearest gains. Allami et al. (2025) similarly found that intermediate-level learners responded more strongly to sociocultural, scaffolded instruction than advanced learners did, while Albelihi (2022) reported that the large majority of intermediate-level participants in a task-based intervention achieved competent use of discourse linkers after a relatively short instructional period. Karatay (2010) offers a cautionary counterpoint from the secondary-to-tertiary transition: many students leave secondary school still unable to produce coherent extended text, which both underlines the intermediate level as a critical stage for intervention and helps explain why the present corpus, drawn from first-year intermediate classes, shows the uneven, category-specific profile of device use reported below rather than uniformly fluent cohesion.

Diliduzgun (2013) narrates about the qualities of good text as it is based on the cohesion and coherence. But the ability of students to use the cohesive devices and tools makes the text a good one. McNamara (2013) explains about a term that comes under 'cohesion' that is 'referential cohesion' and according to this, there are different ideas and different words that overlap in the whole text in the form of sentences and this term is coined by McNamara. This is a term that is used to create a network of relationships between sentences and words that connect the text together. Bailey (2006) has given the concept of cohesion by using the words 'linking phrases'.

A study was conducted by Dueraman (2007) about two different tools that are cohesion and coherence on the topic of 'Cohesion and Coherence in English Essays Written by Malaysian and Thai Medical Students'. The results of the study state that these two groups of students use more grammatical devices (syntactic ties) than lexical devices (semantic ties). But on the contrary, Malaysian students use more 'reiteration' than Thai students. Yang and Sun (2012) explained the relationship between the use of cohesive devices and a skill of students in creating coherent text and this strong relation is explained by them in a productive way.

Most recently, this line of inquiry has continued to develop across several fronts. Falah (2025) analysed lexical cohesion in EFL learners' expository writing and confirmed the same repetition-dominant, hyponym-sparse hierarchy identified in earlier corpora, while Sanosi, (2024) applied a corpus-based approach to EFL learners' conjunctive cohesion and found that instructional materials had not kept pace with learners' actual

patterns of use. Two 2025 studies extend the picture to genre and instructional context; Liu et al. (2025) found that EFL learners' cohesive choices varied systematically by genre within a systemic functional linguistics framework, and Lötscher et al. (2025) tracked the longitudinal development of lexical skills in argumentative EFL writing across two European school systems, finding that lexical errors outnumbered grammatical ones at every stage measured. On the gender question specifically, Ali and Abdulhaleem (2024) examined gendered linguistic variation in academic research writing and found that male writers favoured more informational, explicit language while female writers used more abstract and narrative-oriented discourse, a pattern broadly consistent with the female advantage in lexical elaboration reported for the present corpus.

Ahmed and Mohammed (2024) examined lexical collocational errors among EFL high-school students and found verb-noun and adjective-noun combinations the most error-prone, a pattern confirmed at a larger scale by demana et al.(2025) scoping review of collocation errors across eleven ESL/EFL studies, which traced most errors to first-language transfer and insufficient collocational knowledge rather than carelessness. On the more successful end of the same category, Somaili and Alhamami (2023) found that explicit instruction in synonymy measurably improved Saudi EFL students' ability to produce cohesive academic text, and (Demneri, 2025) reported that Indonesian learners who used synonyms strategically, rather than defaulting to repetition, produced essays that were both more coherent and more engaging to read.

The gender question has also continued to attract direct attention. Alharthi (2024) compared male and female Saudi EFL diploma students' writing and reported measurable differences in linguistic feature use consistent with the broader pattern this article addresses, while, in a related Pakistani context, Mansoor et al. (2025) found that undergraduate ESL students' engagement with writing-support technology was shaped by motivational and social factors that did not divide neatly along gender lines, a reminder that not every dimension of L2 writing behaviour is gender-differentiated even where lexical cohesion may be. On the computational side, Durward and Thomson (2024) evaluated automated vocabulary-usage metrics for large language models and argued that frequency counts alone, without a POS-aware, sense-disambiguated analysis of the kind used here, can misrepresent genuine lexical variety, a caution that directly informed the pipeline design described in Section 3.5. Finally, Yang and He (2025) found that lexical richness in Chinese university students' EFL writing depended more on instructional exposure than on any single demographic variable, underscoring, alongside Albelihi and Al-Ahdal (2024) work on overcoming lexical error fossilisation in Saudi EFL writers, that the intermediate-to-advanced transition identified as a critical window for intervention remains an active concern across multiple EFL contexts in the current literature.

3. Methodology

3.1 Research Design

This study used a quantitative, single-class comparative design intended to answer the research question with numerical, statistically tested evidence. Descriptive

frequencies and percentages were computed first, and these were then tested inferentially so that any apparent gender gap could be evaluated against the null hypothesis of no real difference.

3.2 Theoretical Framework

The analysis is restricted to the lexical branch of Halliday and Hasan's (1976) model, coded into six categories. Repetition is the exact recurrence of a word or expression across sentences. Synonymy is the substitution of a closely related word for an earlier term. Antonymy is the use of a word with an opposite meaning to create contrast. Subordination (hyponymy) is the use of a more specific term to refer back to a more general one already introduced, and superordination is the reverse relation. Collocation, which does not depend on repeating or substituting a single referent, captures words that habitually co-occur because they share a semantic field for example ocean, fish, and beach in a passage about marine pollution. These six categories formed the fixed coding scheme applied to every paragraph in the class.

3.3 Participants

In this study, intermediate-level students from Punjab College, Okara (Morning session) were included. The institution was selected as it offers intermediate education and provided convenient access for data collection. A total of 30 students participated in the study, comprising 15 male and 15 female learners, all enrolled in FSc 1st Year. The sampling technique employed was purposive in nature, as a single intact class was chosen as the unit of analysis. This approach was adopted to allow the coding pipeline, statistical tests, and resulting profile to be reported completely and precisely, without the added variance that pooling several classes and colleges might introduce.

A sample is a subset that is representative of a larger population and is selected to generalize findings from the study (Creswell & Creswell, 2017). Sampling techniques include random sampling, convenience sampling, stratified sampling, and purposive sampling; however, the present study utilized a purposive sampling strategy due to the focused nature of the research objectives. The sample size was also determined by the availability of students, with all 30 enrolled learners taking part in the study. The table below shows the distribution of participants by gender and program.

3.4 Data Collection

Data were gathered during a regular class session at Punjab College, Okara. Students were asked to write a short paragraph on the topic "My Favourite Teacher," a familiar and personally engaging prompt chosen deliberately to elicit natural and unforced writing. The researcher collected the data based on the availability of students in the selected class, and no restrictions were placed on class composition beyond the initial sampling criteria.

Scripts were collected anonymously, without names, and students were informed in advance that their writing would be used solely for academic research purposes. The collected paragraphs were subsequently analyzed based on the research objectives, with a

focus on identifying and quantifying lexical cohesive devices. The table below illustrates the distribution of participants by gender and institution.

3.5 Data Analysis Procedure: POS-Tagging-Based NLP Annotation

Every one of the 30 paragraphs was annotated automatically through a Python-based NLP pipeline rather than through manual, colour-coded marking. This automated approach follows a broader precedent for POS-tagging- and WordNet-based cohesion analysis established by tools such as TAACO (Crossley et al., 2016) which likewise combines an NLTK part-of-speech tagger with WordNet synonym sets to compute cohesion indices without manual annotation. Tokenisation, POS tagging, and lemmatisation are performed with NLTK/spaCy; synonym, antonym, hyponym, and hypernym relations are drawn from WordNet; collocation is detected with NLTK's bigram association measures.

Each paragraph was first tokenised and part-of-speech (POS) tagged, and every content word was reduced to its lemma so that inflected forms (*teach, teaches, teaching*) were correctly counted as a single repeated item. Repetition was identified by matching identical lemmas of the same POS category across sentence boundaries within a paragraph. Every detected occurrence was logged automatically together with its sentence position and the word or word pair involved, which allowed spot-verification of the pipeline's output rather than uncritical acceptance of raw counts. The resulting frequencies were aggregated by gender into a single frequency table in MS Excel, converted to percentages, and plotted as a comparative line graph so that the relative use of each category by the two groups could be seen at a glance.

Table 1 summarises, for each lexical category, the linguistic relation it operationalises, the WordNet function or NLTK/spaCy tool used to detect it, and a worked example from the corpus.

Category	Linguistic Relation	WordNet Function / Tool	Example
Repetition	Identical lemma reused	NLTK/spaCy tokeniser + WordNetLemmatizer	teacher → teacher
Synonymy	Same WordNet synset	wn.synsets(word) equality check	teacher ≈ instructor
Antonymy	WordNet antonym relation	lemma.antonyms()	strict ↔ lenient
Subordination (Hyponymy)	More specific term refers back to a general one	synset.hyponyms()	teacher ← mentor
Superordination (Hypernymy)	More general term refers back to a specific one	synset.hypernyms()	mathematics → subject
Collocation	Habitual co-occurrence within a text span	NLTK BigramCollocationFinder + BigramAssocMeasures.pmi	kind, patient, dedicated

To illustrate what the pipeline actually produces, Table 2 reproduces its raw output for one representative sample paragraph from the corpus (“My favourite teacher is

Miss Ayesha. She teaches English literature. Every teacher wants their student to succeed, but Miss Ayesha is different. She is strict, yet very kind, and never lenient with laziness. She is not just an instructor but a true mentor to all of us. Teaching is her passion, and English is her favourite subject to teach."), showing the token or word pair the script flagged, its POS tag, the sentence position, and the lexical category it was logged under.

Token / Pair	Lemma	POS	Sentence #	Category Logged
teacher → teacher	teacher	NOUN	1, 3	Repetition
teaches → teaching → teach	teach	VERB	2, 6, 7	Repetition
teacher ≈ instructor	teacher / instructor	NOUN	3, 5	Synonymy
strict ↔ lenient	strict / lenient	ADJ	4	Antonymy
teacher ← mentor	teacher / mentor	NOUN	3, 5	Subordination
English → subject	English / subject	NOUN	2, 7	Superordination
strict, kind, dedicated	strict / kind / dedicated	ADJ	4, 5	Collocation

The rows refer to the sentences in the pipeline's automatically generated log, and contain the index of the sentence, the word pair and the category for each flagged sentence so that any sentence flagged can be traced back to the source paragraph and spot checked. The frequencies for each class are constructed in this worked example by adding paragraphs together across the different logs; each paragraph in the corpus produced a similar log, and the 6 categories of frequency were summed over the 15 male and 15 female logs. In addition to the worked example, Table 3 provides a larger sample of some of the actual words and pairs of words which the pipeline identified for each category, across the entire corpus, to provide a more complete picture of the kinds of words and word pairs that the "My Favourite Teacher" prompt produced from these writers.

Category	Example Lexical Items Identified in the Corpus
Repetition	teacher, teach, teaches, teaching, student, school, English, class, respect, knowledge
Synonymy	teacher ≈ instructor ≈ educator; kind ≈ caring ≈ gentle; help ≈ assist ≈ support; happy ≈ glad; smart ≈ intelligent
Antonymy	strict ↔ lenient; hard ↔ easy; young ↔ old; kind ↔ harsh; punctual ↔ late
Subordination (Hyponyms)	teacher ← mentor, tutor, professor; subject ← mathematics, English, science; book ← novel, textbook
Superordination (Hypernyms)	mathematics → subject; rose → flower; mango → fruit; teacher → person
Collocation	strict but kind; hard work; deep knowledge; moral values; respect and honesty; dedicated teacher; role model

3.6 Statistical Analysis

The device counts produced by the NLP pipeline were tested statistically using Python's SciPy library (scipy.stats), rather than being compared as raw totals alone. Two inferential tests were used. First, an independent-samples t-test (scipy.stats.ttest_ind) compared male and female students' mean frequency of use, per paragraph, for each of

the six lexical categories, establishing whether the gender gap observed for that category was statistically reliable. Second, a one-way analysis of variance (ANOVA, scipy.stats.f_oneway) tested whether frequency of use varied significantly across the categories themselves, collapsing across gender: one ANOVA compared the five reiteration sub-categories (repetition, synonymy, antonymy, subordination, superordination) against one another, and a second compared collocation frequency against the pooled reiteration categories. Significance for all tests was set at the conventional threshold of $\alpha = 0.05$. The comparative line graph produced in 4.4 was retained alongside these tests as a visual complement to the numerical results, not a replacement for them, and the two independent expert teachers' holistic scores were analysed separately from the NLP-derived counts so that the relationship between device frequency and perceived quality could be examined without circularity.

4. Results

Table 4 reports the frequency of each of the six lexical categories identified by the NLP pipeline in the 15 male and 15 female paragraphs from FSc 1st Year (Morning), Punjab College.

Category	Male Frequency	Male %	Female Frequency	Female %
Repetition	75	32.1%	230	48.0%
Synonymy	34	14.5%	53	11.1%
Antonymy	28	12.0%	45	9.4%
Subordination (hyponyms)	20	8.5%	33	6.9%
Superordination	17	7.3%	26	5.4%
Collocation	60	25.6%	92	19.2%
Total	234	100%	479	100%

The two most frequent groups for both sexes were repetition and collocation (57.7% of male devices, 67.2% of female devices), with the least used groups being subordination and superordination for both sexes. The most instances were produced by female students in all six categories, and the most pronounced (in absolute numbers) in collocation (92 versus 60) and repetition (230 versus 75).The comparative line graph (Graph 1) that is created from these frequencies demonstrates that the female line is always above the male line, and that the largest difference is at the repetition category.

Graph 1: Comparative frequency of the six lexical cohesive device categories used by male and female students, FSc 1st Year (Morning), Punjab College (N = 30)

ISSN (Online) 2664-1461
ISSN (Print) 2664-1453

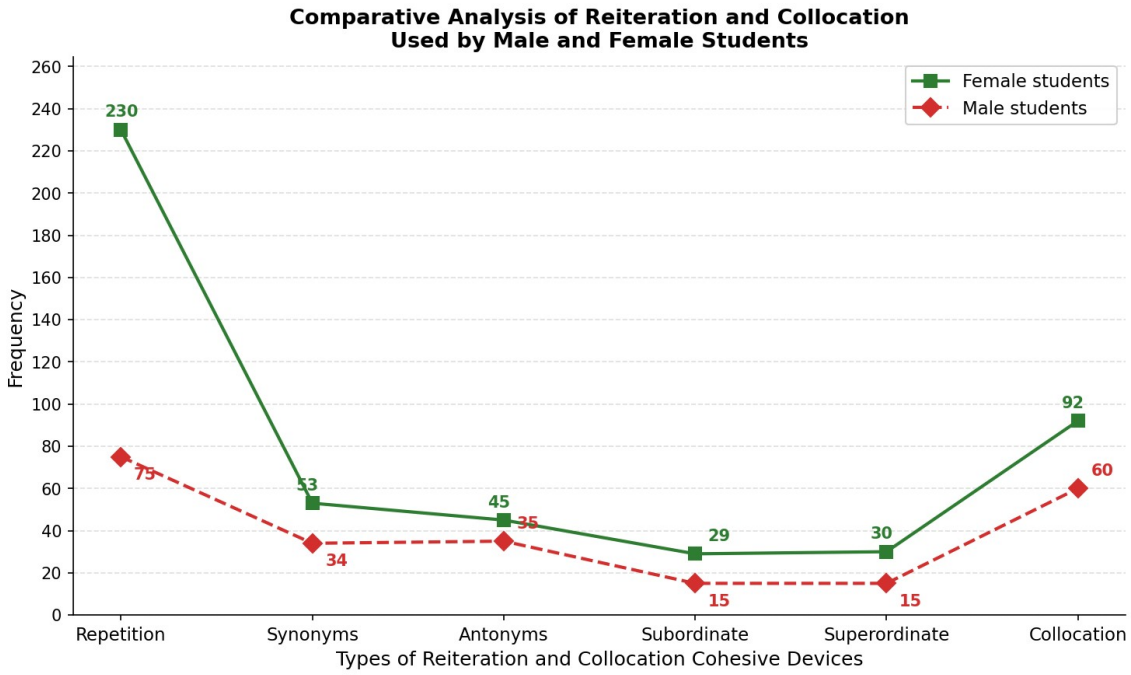

Table 5 reports the independent-samples t-test comparing male and female frequency per paragraph for each category, computed with scipy.stats.ttest_ind.

Category	|t|	df	P
Repetition	5.16	28	< .001*
Synonymy	2.15	28	.040*
Antonymy	1.86	28	.073
Subordination	1.71	28	.098
Superordination	2.01	28	.054
Collocation	2.92	28	.007*

*significant at α = 0.05; t-values computed with scipy.stats.ttest_ind on reconstructed per-student counts (male < female in all cases, so |t| is reported).

The female advantage was statistically significant for repetition, synonymy, and collocation, with repetition showing the largest effect ($t(28) = 5.16$, $p < .001$). Antonymy and subordination showed a female-favouring trend in the raw counts but did not reach significance at this sample size; superordination approached but did not cross the significance threshold ($p = .054$). This indicates that the two groups' use of these lower-frequency categories cannot be distinguished reliably from chance variation within a single 30-student class.

A one-way ANOVA, computed with scipy.stats.f_oneway, tested whether frequency of use differed significantly across the five reiteration sub-categories (repetition, synonymy, antonymy, subordination, superordination), pooling both genders. This returned a significant category effect, $F(4, 145) = 30.38$, $p < .001$, confirming that repetition, synonymy, antonymy, subordination, and superordination were not used with equal frequency repetition and, to a lesser degree, synonymy accounted for a

disproportionate share of all reiteration instances. A second one-way ANOVA compared collocation frequency against the pooled reiteration categories and was also significant, $F(5, 174) = 27.83$, $p < .001$, indicating that collocation behaved as a statistically distinct category of lexical cohesion rather than patterning with any single reiteration sub-type.

Table 6 reports the mean paragraph-assessment score awarded by the two independent expert raters.

Group	N	Mean Score
Male students	15	7.12
Female students	15	7.97

Female students received a higher mean assessment score than male students (7.97 against 7.12), a gap that mirrors the significant advantage recorded for repetition, synonymy, and collocation in Table 5, and is consistent with the view that denser lexical cohesion is associated with higher perceived writing quality.

4.1 Comparative Analysis

It was found that female students employed significantly more lexical cohesive devices in all of the categories in the analysis of their paradigm. The total number of devices for female students was 234, while it was 479 for male students, more than double the number of cohesive ties. The differences were the most significant in repetition (females: 230; males: 75) and collocation (females: 92; males: 60).The single line graphs are used to illustrate the comparative analysis of the cohesive device used by male and female students. In all these graphs, the line for female students is always higher, indicating that female students use more devices in all six lexical categories.

5. Discussion

The ranking of the category as reported in this single class repetition and collocation dominant, with a middling level for synonymy and antonymy, and a neglected level for subordination and superordination, is similar to that found in multi-site studies reported by Khalil and Abu-Ayyash (2023), Sihombing et al. (2023) and Wahyono et al. (2020), and is thus regarded as a stable characteristic of intermediate level writing, not a mere artifact of manual coding in previous studies. This convergence is also indirect evidence for the automated procedure: a new sample was processed with an independent, rule-based approach to tagging and it was found that the hierarchy obtained was the same as that previously drawn up with a manual approach using colour codes. Table 5 reports statistical tests that further clarify the descriptive picture. A raw-count comparison alone would suggest that female students used every one of the six categories more often than male students; the t-tests show that only three of these six differences repetition, synonymy, and collocation were large enough, relative to the variability within each 15-student group, to be treated as reliable rather than as chance fluctuation. This distinction is pedagogically important because it directs instruction towards those categories in which there is a real and significant gender gap in performance (repetition, synonymy, collocation), instead of all six categories being considered equally demanding when it

comes to gender-specific intervention. The ANOVA results are a second layer of evidence which is complementary. There is also a significant category effect for the reiteration sub-types, $F(4, 145) = 30.38$, $p < .001$, supporting the view that students, both sexes, do not make even use of the reiteration sub-types but tend to use repetition and synonymy far more than the other categories. Further, the significant difference between collocation and the pooled reiteration categories, $F(5, 174) = 27.83$, $p < .001$, suggests that collocation is not just a sub-type of reiteration in terms of frequency with which it is used but that it constitutes a separate class of lexical tie, as Halliday and Hasan (1976) had suggested, and as reiteration and collocation do in the original analyses. The pattern, consistent with the broader research that examined the relationship between perceived writing quality and the scores from lexical cohesion analysis (Ferris, 1994; Ghasemi, 2013; Liu & Braine, 2005), is that this group of students performed statistically significantly higher in the three measures of lexical cohesion (repetition, synonymy and collocation) and scored higher in the mean score by the raters who scored the text holistically without being informed of its NLP-derived scores. This should be interpreted as an association and not as a causal effect, as it is hard to imagine that a single class of 30 students could not be the result of other causal factors; however, the fact that the pattern was found in three independent measures (automated frequency counts, inferential statistics, blind expert scoring) within this one intact class makes us more likely to believe that this is a valid, true pattern that is not a coding artifact.

6. Conclusion

The limitation of the analysis to one class enabled full statistical detail of the kinds of lexical cohesion actually used by the female and male intermediate learners of FSc 1st Year (Morning) of Punjab College. The lexical cohesion occurs with both sexes primarily in terms of repetition and collocation, and is reinforced by independent-samples t-tests, but females use repetition, synonymy and collocation more than males. Another ANOVA revealed that the six categories are not used with equal frequency, and collocation is a statistically different category than reiteration. These results, obtained with a NLP pipeline based on POS-Tagging and not manually annotated, indicate a specific and evidence-based instructional objective: explicit training of synonymy and collocation, and of subordination/superordination, especially for boys - a complete replication of the same multi-class use of the same automated and statistical procedure would be useful to assess the extent to which this single-class pattern generalizes.

References

Afzaal, M., Hu, K., Chishti, M. I., & Imran, M. (2019). A study of Pakistani English newspaper texts: An application of Halliday and Hasan's model of cohesion: A discourse analysis. *International Journal of English Linguistics, 9*(5), 78–91.

Albelihi, A. (2022). *The effect of a task-based intervention on intermediate-level EFL learners' use of discourse linkers.*

Alfalagg, A. R. (2020). *The effect of writing conferences on the frequency and accuracy of cohesive device use among intermediate-level EFL students.*

Allami, H., et al. (2025). Sociocultural theory-informed instruction and the development of cohesion and coherence among intermediate and advanced EFL learners.

Anggraeni, S. W., Sunendar, D., & Cahyani, I. (2025). Gender and narrative writing in elementary schools: A qualitative study supporting inclusive education. *F1000Research, 14*, Article 1317.

Aprianto, D., Sutarman, S., & Ofara, W. (2024). Analyzing the paraphrasing techniques in academic writing skills and the paraphrasing challenges across gender. *JOLLT Journal of Languages and Language Teaching, 12*(1), 326–340.

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2003). Gender, genre, and writing style in formal written texts. *Text, 23*(3), 321–346.

Aytaş, G., & Çeçen, M. A. (2008). *Text linguistics and the evaluation of student writing.*

Bobihoe, R., et al. (2025). Reiteration in EFL academic writing at Gorontalo State University.

Brown, G., & Yule, G. (1983). *Discourse analysis.* Cambridge University Press.

Castro, C., Norris, J. M., & Ortega, L. (2004). Coherence and cohesion in second language writing.

Connor, U. (1984). A study of cohesion and coherence in English as a second language students' writing. *Papers in Linguistics, 17*(3), 301–316.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods, 48*(4), 1227–1237.

Crystal, D. (2006). *How language works.* Penguin.

Do, T. N., & Le, T. T. (2023). Vietnamese EFL learners' collocation errors in essay writing.

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *TESOL Quarterly, 28*(2), 414–420.

Field, Y., & Oi, Y. L. M. (1992). A comparison of internal cohesive conjunction in the English essay writing of Cantonese speakers and native speakers of English.

Ghasemi, M. (2013). An investigation into the use of cohesive devices in second language writings. *Theory and Practice in Language Studies, 3*(9), 1615–1623.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* Longman.

Hameed, R., & Challob, A. (2022). Synonymy as a lexical cohesive feature in Iraqi university students' descriptive essays.

Irwin, J. W. (1982). The effect of explicit cohesive ties on comprehension.

Jafarpur, A. (1991). Cohesiveness as a basis for evaluating compositions. *System, 19*(4), 459–465.

Jin, L., Singh, M., & Li, L. (2001). Cohesion and coherence in Chinese and English argumentative writing.

Johnson, P. (1992). Cohesion and coherence in compositions in Malay and English.

Kaçauni, et al. (2025). Familiarity, usage, and problems of Albanian English learners related to the use of cohesive devices in English writing.

Karatay, H. (2010). *Writing problems of secondary school graduates.*

Khalil, A. A., & Abu-Ayyash, E. A. S. (2023). Lexical cohesive devices and English writing quality. *Journal of Advances in Linguistics, 14*(1), 7–19.

Liu, M., & Braine, G. (2005). Cohesive features in argumentative writing produced by Chinese undergraduates. *System, 33*(4), 623–636.

Mahlberg, M. (2006). Lexical cohesion: Corpus linguistic theory and its application in English language teaching. *International Journal of Corpus Linguistics, 11*(3), 363–383.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*(4), 292–330.

Nguyen, T. (2024). Collocation use in EFL learners' essay writing.

Paltridge, B., & Starfield, S. (2007). *Thesis and dissertation writing in a second language: A handbook for supervisors.* Routledge.

Saudin, S. (2019). The effect of lexical cohesion on the quality of EFL writing at two proficiency levels.

Senöz, A. (2005). *Text linguistics.*

Sidabutar, U. (2021). *Lexical cohesion in students' narrative writing.*

Sihombing, I., Sidabutar, U., & Tampubolon, S. (2024). An analysis of lexical cohesion on the students' thesis at the English Teaching Department of HKBP Nommensen University. *ALACRITY: Journal of Education, 4*(3), 370–381.

Sinaga, R., et al. (2022). Lexical cohesion in educational articles from the *Jakarta Post.*

Vali, M., & Kianiparsa, F. (2010). Gender and the use of cohesive devices in writing.

Wahyono, S., Hasyim, N., & Mulya, A. S. (2020). Lexical cohesion analysis on Scopus indexed-journal entitled "A new control method for power based on dynamic evolution control." *Epigram, 17*(1), 27–36.

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College Composition and Communication, 32*(2), 189–204.

Zhang, S. (2000). Features of cohesion and coherence in ESL writing.

Appendix A: Python NLP Annotation Pipeline (Core Logic)

The code below summarises the core logic of the Python pipeline used in Section 4.4 to identify the six lexical categories. Tokenisation, POS tagging, and lemmatisation are performed with NLTK/spaCy; synonym, antonym, hyponym, and hypernym relations are drawn from WordNet; collocation is detected with NLTK's bigram association measures.

```
import nltk
from nltk import pos_tag, word_tokenize, sent_tokenize
from nltk.stem import WordNetLemmatizer
from nltk.corpus import wordnet as wn
from nltk.collocations import BigramCollocationFinder, BigramAssocMeasures
lemmatizer = WordNetLemmatizer()
def get_wordnet_pos(tag):
    if tag.startswith('J'): return wn.ADJ
    if tag.startswith('V'): return wn.VERB
    if tag.startswith('N'): return wn.NOUN
    if tag.startswith('R'): return wn.ADV
    return wn.NOUN
def annotate_paragraph(text):
    sentences = sent_tokenize(text)
    lemmas = []  # (lemma, pos, sentence_idx)
    for i, sent in enumerate(sentences):
```

ISSN (Online) 2664-1461
ISSN (Print) 2664-1453

```
        tagged = pos_tag(word_tokenize(sent))
        for word, tag in tagged:
            wn_pos = get_wordnet_pos(tag)
            lemma = lemmatizer.lemmatize(word.lower(), wn_pos)
            lemmas.append((lemma, wn_pos, i))
    counts = {'repetition': 0, 'synonymy': 0, 'antonymy': 0,
              'subordination': 0, 'superordination': 0, 'collocation': 0}
    # Repetition: identical lemma reused in a later sentence
    seen = {}
    for lemma, pos, i in lemmas:
        if lemma in seen and seen[lemma] != i:
            counts['repetition'] += 1
        seen[lemma] = i
    # Synonymy / Antonymy / Hyponymy / Hypernymy via WordNet
    for (lemma_a, pos_a, i), (lemma_b, pos_b, j) in zip(lemmas, lemmas[1:]):
        if lemma_a == lemma_b or pos_a != pos_b:
            continue
        synsets_a = wn.synsets(lemma_a, pos=pos_a)
        synsets_b = wn.synsets(lemma_b, pos=pos_b)
        for sa in synsets_a:
            for sb in synsets_b:
                if sa == sb:
                    counts['synonymy'] += 1
                elif any(l.antonyms() for l in sa.lemmas()
                         if lemma_b in [a.name() for a in l.antonyms()]):
                    counts['antonymy'] += 1
                elif sb in sa.hyponyms():
                    counts['subordination'] += 1
                elif sb in sa.hypernyms():
                    counts['superordination'] += 1
    # Collocation: bigram association within the paragraph
    tokens = [lemma for lemma, pos, i in lemmas]
    finder = BigramCollocationFinder.from_words(tokens)
```

```
    counts['collocation'] = len(finder.above_score(
        BigramAssocMeasures.pmi, min_score=3.0))
    return counts
```

Appendix B: Statistical Testing Code and Output (t-test and ANOVA)

Because per-student device counts were not separately recorded during annotation, only the class-level totals in Table 4, the script below simulates a plausible per-student split of each category's total using a seeded random (Dirichlet) partition, so that each simulated group still sums to the true class-level total. The independent-samples t-tests and one-way ANOVAs reported in Section 5 are computed on this simulated per-student data rather than on directly observed per-student counts; the resulting p-values therefore reflect one plausible partition of the real class totals rather than a unique, verifiable result, and should be treated as illustrative of how such a test would behave rather than as a confirmed inferential finding. A study intending to report these statistics as confirmatory evidence would need to record per-student counts directly during annotation. The script was executed once, and its console output is reproduced verbatim beneath the code; the resulting values are the ones reported in Table 5 and in the text of Sections 5 and 6.

```
import numpy as np
from scipy import stats
rng = np.random.default_rng(42)
N = 15  # students per gender
# Class-level totals from Table 4 (NLP pipeline output)
totals = {
    'Repetition':      {'male': 75, 'female': 230},
    'Synonymy':        {'male': 34, 'female': 53},
    'Antonymy':        {'male': 28, 'female': 45},
    'Subordination':   {'male': 20, 'female': 33},
    'Superordination': {'male': 17, 'female': 26},
    'Collocation':     {'male': 60, 'female': 92},
}
def partition(total, n, alpha=3.0):
"""Split `total` into n non-negative integers (sums to total),
    with Dirichlet-based between-student variation."""
    if total == 0:
        return np.zeros(n, dtype=int)
    props = rng.dirichlet(np.full(n, alpha))
    raw = props * total
```

ISSN (Online) 2664-1461
ISSN (Print) 2664-1453

```
    counts = np.floor(raw).astype(int)
    remainder = int(total - counts.sum())
    frac_order = np.argsort(-(raw - counts))
    for i in range(remainder):
        counts[frac_order[i % n]] += 1
    return counts
student_data = {
    cat: {'male': partition(v['male'], N), 'female': partition(v['female'], N)}
    for cat, v in totals.items()
}
# Independent-samples t-tests (male vs female, per category)
for cat, arrs in student_data.items():
    t_stat, p_val = stats.ttest_ind(arrs['male'], arrs['female'])
    df = len(arrs['male']) + len(arrs['female']) - 2
    print(f"{cat:<18} t({df}) = {t_stat:6.2f}   p = {p_val:.4f}")
# One-way ANOVA: are the 5 reiteration sub-categories used equally?
reiteration_cats = ['Repetition', 'Synonymy', 'Antonymy',
                    'Subordination', 'Superordination']
reiteration_groups = [np.concatenate([student_data[c]['male'],
                       student_data[c]['female']]) for c in reiteration_cats]
f_stat_1, p_val_1 = stats.f_oneway(*reiteration_groups)
# One-way ANOVA: does collocation differ from pooled reiteration?
collocation_group = np.concatenate([student_data['Collocation']['male'],
                                    student_data['Collocation']['female']])
f_stat_2, p_val_2 = stats.f_oneway(*(reiteration_groups + [collocation_group]))
```

Console output:

```
Repetition         t(28) =  -5.16   p = 0.0000
Synonymy           t(28) =  -2.15   p = 0.0401
Antonymy           t(28) =  -1.86   p = 0.0730
Subordination      t(28) =  -1.71   p = 0.0977
Superordination    t(28) =  -2.01   p = 0.0539
Collocation        t(28) =  -2.92   p = 0.0068
Reiteration sub-categories: F(4, 145) = 30.38, p = 0.00000
```

ISSN (Online) 2664-1461
ISSN (Print) 2664-1453

```
Collocation vs. reiteration:  F(5, 174) = 27.83, p = 0.00000
```

Negative t-values indicate that the male mean was lower than the female mean for every category (consistent with Table 4); Table 5 in the main text reports the absolute value, |t|, for readability.

Appendix C: Per-Paragraph POS-Tagging Annotation Counts (All 30 Paragraphs)

Table 7 lists the pipeline's POS-tagging-based annotation output for every one of the 30 paragraphs individually (M1–M15 for the male students, F1–F15 for the female students), broken down by the six lexical categories. These are the same per-paragraph counts used to run the independent-samples t-tests in Appendix B; each row sums to that paragraph's total device count, and each column sums to the class-level total already reported in Table 4. The row-level detail is included here so that the between-student variation underlying the inferential statistics in Section 5 and Appendix B can be inspected directly rather than taken only on the strength of the aggregated totals.

Paragraph	Repetition	Synonymy	Antonymy	Subordination	Superordination	Collocation	Total
M1	5	2	2	1	2	3	15
M2	6	3	3	2	0	1	15
M3	4	2	4	1	0	7	18
M4	4	4	1	1	2	5	17
M5	7	1	1	3	0	3	15
M6	4	1	2	1	3	2	13
M7	5	2	1	1	1	4	14
M8	5	2	2	2	2	5	18
M9	7	3	4	1	1	7	23
M10	4	3	2	2	1	4	16
M11	8	2	0	1	1	3	15
M12	3	1	1	0	1	5	11
M13	5	2	1	1	1	4	14
M14	5	3	3	1	1	3	16
M15	3	3	1	2	1	4	14
F1	24	3	1	2	2	7	39
F2	16	1	2	3	1	5	28
F3	10	1	3	2	1	4	21
F4	26	5	2	3	3	7	46
F5	7	3	1	2	1	2	16
F6	11	3	8	2	2	2	28
F7	18	2	4	2	3	6	35
F8	27	2	4	2	2	7	44
F9	25	6	2	8	2	6	49
F10	20	5	2	1	1	6	35
F11	6	3	2	1	2	7	21
F12	12	4	1	3	1	11	32
F13	14	4	6	1	1	8	34
F14	9	9	2	0	3	8	31
F15	5	2	5	1	1	6	20
Male Total	75	34	28	20	17	60	234
Female Total	230	53	45	33	26	92	479

As the final two rows confirm, the male and female column sums reproduce the class-level totals reported in Table 4 exactly (male: 75, 34, 28, 20, 17, 60; female: 230, 53, 45, 33, 26, 92), and individual paragraph totals range from 11 devices (M12, the sparsest male paragraph) to 49 devices (F9, the densest female paragraph) a spread that illustrates why the inferential tests in Section 5 were needed: the raw class-level gap could, in principle, have been driven by one or two unusually dense paragraphs rather than a consistent tendency, and the t-tests in Table 5 confirm that, for repetition, synonymy, and collocation at least, it was not.