

Lexical profile of academic written texts produced by EFL students at the Technical University of Cotopaxi

Perfil léxico de textos académicos escritos producidos por estudiantes de inglés como lengua extranjera en la Universidad Técnica de Cotopaxi

- ¹ Jessica Dayanna Granja Rubio  <https://orcid.org/0009-0008-6530-3586>
Universidad Técnica de Cotopaxi (UTC), Latacunga, Ecuador.
Maestría en Pedagogía del Inglés
jessica.granja7575@utc.edu.ec
- ² Olga Lorena González Ortiz  <https://orcid.org/0000-0003-0064-1601>
Universidad Técnica de Cotopaxi (UTC), Latacunga, Ecuador.
olga.gonzalez@utc.edu.ec



Artículo de Investigación Científica y Tecnológica

Enviado: 15/07/2026

Revisado: 14/08/2026

Aceptado: 02/10/2026

Publicado: 21/10/2026

DOI: <https://doi.org/10.33262/exploradordigital.v10i4.3789>

Cítese:

Granja Rubio, J. D., & González Ortiz, O. L. (2026). Perfil léxico de textos académicos escritos producidos por estudiantes de inglés como lengua extranjera en la Universidad Técnica de Cotopaxi. Explorador Digital, 10(4), 120-135.
<https://doi.org/10.33262/exploradordigital.v10i4.3789>



EXPLORADOR DIGITAL, es una Revista electrónica, **Trimestral**, que se publicará en soporte electrónico tiene como **misión** contribuir a la formación de profesionales competentes con visión humanística y crítica que sean capaces de exponer sus resultados investigativos y científicos en la misma medida que se promueva mediante su intervención cambios positivos en la sociedad. <https://exploradordigital.org>
La revista es editada por la Editorial Ciencia Digital (Editorial de prestigio registrada en la Cámara Ecuatoriana de Libro con No de Afiliación 663) www.celibro.org.ec



Esta revista está protegida bajo una licencia Creative Commons en la 4.0 International. Copia de la licencia: <http://creativecommons.org/licenses/by-nc-sa/4.0/>



Palabras claves:

Bloques léxicos;
diversidad léxica;
escritura
académica;
lingüística de
corpus,
perfil léxico.

Resumen

Introducción. Aunque la escritura académica exige precisión y sofisticación, los estudiantes de lenguas extranjeras suelen presentar limitaciones en su repertorio productivo, lo que evidencia la necesidad de evaluar su competencia léxica. **Objetivo.** Determinar el perfil léxico de ensayos académicos producidos por estudiantes de inglés como lengua extranjera (EFL) en la Universidad Técnica de Cotopaxi (UTC), Ecuador. **Metodología.** Bajo un paradigma positivista, se analizó un corpus de 43 ensayos académicos de la carrera de Pedagogía de Idiomas. Mediante herramientas digitales (VP y AntConc), se midió la diversidad, densidad y frecuencia léxica. **Resultados.** Los resultados revelaron una maduración estadísticamente significativa entre cohortes ($d = 0.72$), incrementando el TTR de 0.48 a 0.56, la densidad léxica de 42.1% a 46.8% y la integración de AWL de 11.5% a 18.8%. No obstante, persistió una marcada dependencia del vocabulario de alta frecuencia (81.2% en sexto semestre), situándose por debajo de los estándares del nivel B2 del MCER (> 0.60 TTR y $> 25\%$ AWL). **Conclusión.** Se concluye que, aunque los estudiantes dominan la macroestructura del ensayo, enfrentan un estancamiento léxico en su repertorio productivo formal, requiriéndose una transición pedagógica hacia la enseñanza explícita de colocaciones académicas y bloques léxicos. **Área de estudio general:** Educación. **Área de estudio específica:** Formación docente. **Tipo de artículo:** original.

Keywords:

Lexical chunks;
lexical diversity;
academic
writing;
corpus
linguistics;
lexical profile.

Abstract

Introduction. Although academic writing demands precision and sophistication, foreign language students often exhibit limitations in their productive repertoire; this highlights the need to assess their lexical competence. **Objective.** To determine the lexical profile of academic essays produced by English as a Foreign Language (EFL) students at the Technical University of Cotopaxi (UTC), Ecuador. **Methodology.** Using a positivist paradigm, a corpus of 43 academic essays from the Language Pedagogy program was analyzed. Utilizing digital tools (VP and AntConc), lexical diversity, density, and frequency were measured. **Results.** The results revealed statistically significant maturation between cohorts ($d = 0.72$), with increases in TTR (from 0.48 to 0.56), lexical density (from 42.1% to 46.8%), and AWL integration (from 11.5%

to 18.8%). However, a marked reliance on high-frequency vocabulary persisted (81.2% in the sixth semester), falling below CEFR B2 standards (> 0.60 TTR and $> 25\%$ AWL). **Conclusion.** It is concluded that, although students have mastered the essay's macro-structure, they face a lexical plateau in their formal productive repertoire, necessitating a pedagogical shift toward the explicit teaching of academic collocations and lexical chunks. **General field of study:** Education. **Specific field of study:** Teacher training. **Article type:** Original.

1. Introduction

The present research is justified by the critical need to understand how students of the Pedagogy of National and Foreign Languages major at the Technical University of Cotopaxi (UTC) utilize vocabulary in academic contexts. This study aims to expand knowledge about the acquisition of academic vocabulary in students of English as a foreign language (EFL) in the Ecuadorian highlands. By using Lexical Profiling metrics, the project will contribute to the field of corpus linguistics by determining the levels of lexical sophistication and diversity in university students. The project demonstrates a novel use of computational tools (such as LexTutor or AntConc) to provide an objective, quantitative analysis of written production, moving away from subjective grading. Kyle et al. (2022) explains that automatic tools analyze texts using predefined linguistic algorithms, which means the analysis is not influenced by teacher or researcher subjectivity. From a macro perspective, lexical mastery is the strongest predictor of academic success in a second language (L2). According to Schmitt (2010) lexical mastery has been identified as one of the strongest predictors of academic success, particularly in academic writing contexts. In the meso context (Latin America), there is a documented gap between curriculum expectations and actual production. Locally, at the micro level (specifically at UTC), there is a lack of systematic studies that verify if the students' vocabulary aligns with the B2 level required by the Common European Framework of Reference (CEFR). At this micro level, the study provides faculty with empirical data to adjust syllabi and teaching strategies. Professionally, a robust lexical profile is essential for students to succeed in international certification exams and their future roles as English educators. As Read (2010) notes “Vocabulary knowledge plays a central role in language assessment and is a major component of proficiency testing.”

Despite sustained curricular efforts in EFL instruction, research has consistently shown that many university students rely excessively on high-frequency vocabulary when producing academic texts, particularly words from the first 1,000–2,000 frequency bands.

According to Laufer (2013) lexical richness and sophistication are key indicators of advanced language proficiency, yet L2 learners' writing often shows limited use of low-frequency and academic vocabulary, which restricts the depth and precision of their expression. This tendency has been widely documented in studies on academic language use. Dang & Webb (2014) point out that academic contexts require vocabulary beyond basic high-frequency words, as these are insufficient to convey complex ideas, discipline-specific concepts, and formal academic meanings. When learners fail to incorporate academic and technical lexical items, their texts lack the linguistic resources needed to meet academic standards. Furthermore, Lu & Hu (2022) explains that limited lexical sophistication directly affects writing quality, arguing that learners who depend on common vocabulary struggle to construct well-developed arguments and to express nuanced meanings. As a result, their written production may appear simplistic, repetitive, and less persuasive, which undermines the overall quality of academic discourse. This lexical limitation also impacts students' academic credibility and professional communicative competence. As Nation (2018) states, vocabulary knowledge is central to communicative competence and is essential for effective participation in academic and professional contexts. Inadequate academic vocabulary not only weakens written arguments but also hinders students' ability to perform successfully in advanced academic tasks and professional roles, particularly for future English educators who are expected to model high-level language use.

This research seeks to answer the following question: *"What is the lexical profile of academic texts written by students of English as a Foreign Language (EFL) at the Technical University of Cotopaxi?"*

To this end, the study aims to:

- Determine the lexical diversity and density of the academic essays written by the students.
- Compare the lexical production of fifth-semester students with that of sixth-semester students of the PINE major.
- Identify the students' lexical level considering the CEFR guidelines.

1.1. Lexical profile and lexical richness in higher education

The lexical profile of a text serves as a diagnostic tool to evaluate a learner's productive vocabulary. In the university context, this construct transcends mere word counts, focusing instead on lexical diversity and sophistication.

Recent research by Fredrick & Craven (2025) emphasizes that metrics such as the Type-Token Ratio (TTR) and the Measure of Textual Lexical Diversity (MTLD) are critical indicators for distinguishing the quality of academic essays. High-performing students tend to use a greater proportion of "low-frequency" words, which characterizes the academic register. Studies conducted at regional institutions, including the UTC (Carvajal et al., 2025), indicate that EFL students in Pedagogy programs often face a "lexical plateau." Their writing frequently relies on high-frequency general English words, which

limits their ability to produce the "lexical density" required for scientific or academic discourse. Tovar (2017) asserts that lexical knowledge has a stronger correlation with the overall quality of written production than grammatical accuracy, suggesting that a rich vocabulary is the primary engine for cohesion and coherence in argumentative writing.

1.2. The Common European Framework of Reference (CEFR) and lexical competence

The CEFR (updated in the 2020 Companion Volume) provides standardized descriptors for categorizing language proficiency. For UTC students, the target is usually the transition from B1 (Threshold) to B2 (Vantage). According to the English Profile Programme (updated by Uchida & Negishi, 2025), the shift toward B2 and C1 levels is marked using specific "criterial features," such as academic collocations and complex discourse markers. Research suggests that many EFL students in Latin American contexts struggle to move beyond the A2 level in writing due to a lack of "lexical breadth." This phenomenon, documented by Carvajal et al. (2025) highlights a discrepancy between the curriculum expectations and the students' actual lexical profiles in academic tasks.

Rather than assuming a linear or official equivalence within the qualitative manuals of the CEFR, the benchmarking of lexical levels relies on the criterial features empirically established by Uchida & Negishi (2025) using generative AI models for CEFR-J text classification, alongside the lexical sophistication and richness metrics indexed in Laufer's (2013) computational profiles through the XS_{lex} estimator. According to these frameworks, the threshold of a consolidated independent user (B2) in argumentative writing tasks is characterized by a high-frequency general vocabulary (K1 and K2) distribution of under 75%, and an integration of low-frequency or Academic Word List (AWL) tokens exceeding 25%.

1.3. Academic writing and production

Academic writing is a cognitively demanding process that requires the integration of linguistic, strategic, and sociolinguistic competences.

Conti (2025) argues that academic writing proficiency is not built on isolated words but on "lexical chunks" or formulaic sequences. Students who master these blocks of meaning demonstrate higher levels of fluency and idiomaticity in their written texts. In the current educational landscape, the comparison between human-produced and AI-assisted texts has become a central theme. Etaat (2026) found that while AI can enhance the lexical sophistication of a text, student-produced texts often lack the variety of "stance markers" (words that express the author's opinion) seen in expert academic writing. Lyashevskaya et al. (2022) demonstrate that as the complexity of the academic task increases, so does the demand on the student's lexical repertoire. For UTC students, this means that writing a research report requires a significantly different lexical profile than writing a descriptive essay.

The phenomenon known as the 'Lexical Plateau' remains a persistent structural challenge in higher education across the region. While local studies at UTC by Carvajal et al. (2025) document this localized limitation, these findings align closely with recent research

spanning the broader Andean region (2023–2025). Investigations in equivalent university environments show that EFL written production frequently fossilizes within high-frequency general English bands (K1 and K2), which hinders a smooth transition toward independent B2 descriptors. Furthermore, an exhaustive literature review across the Scielo and Redalyc databases reveals a pronounced lack of empirical studies measuring written lexical profiles under B1–B2 constructs in Ecuador, positioning this study as a pioneering reference point for university linguistic profiling in the country.

2. Methodology

- Epistemological paradigm: positivist. The research relies on objective measurement of linguistic data and statistical analysis of word frequencies to obtain verifiable and generalizable results. As Park et al. (2020) state, the positivist paradigm is based on the belief that reality is objective, observable, and measurable, and that knowledge can be obtained through systematic empirical evidence.
- Research approach: quantitative. The study focuses on collecting numerical data (word counts, lexical richness indices) processed through specialized corpus analysis software (Creswell & Creswell, 2017).
- Research design: non-experimental and cross-sectional. Texts were collected at a single point in time without manipulating the variables (Hunziker & Blankenagel, 2021).
- Research type: descriptive with an explanatory transition. It systematically describes characteristics of vocabulary use as they exist without establishing direct causal relationships (Siedlecki, 2020).

To transition from a strictly descriptive scope toward a descriptive-explanatory design, this study incorporated a qualitative triangulation method based on Key Word in Context (KWIC) concordance lines combined with an outlier analysis. Specifically, five outlier cases were isolated (e.g., students M. Q. from 5th 'B' and D. L. from 6th 'A') whose compositions defied classroom averages by registering over 25% of their vocabulary within the advanced AWL range. These subjects participated in a brief stimulated-recall qualitative interview designed to explore their exogenous learning variables, such as independent reading habits, exposure to authentic target language corpora, and autonomous use of digital tools. This step provided a phenomenological explanation for the variations observed in their lexical sophistication profiles.

2.1. Participants and Sampling

For the selection of participants, a non-probability convenience sampling design was employed, justified by direct administrative access and institutional availability. Specifically, students from the fifth (5th) and sixth (6th) semesters of the PINE major at UTC were selected due to their strategic placement within the macro-curricular structure. The fifth-semester students were chosen because, according to the curriculum, they are currently taking subjects such as "Writing Skill Practice," while in the sixth semester they

have already completed that subject and are taking subjects such as project design. This curricular transition provides an ideal cross-sectional framework to verify the evolution of the lexical profile.

2.2. Variable operationalization

Table 1

Variable Operationalization Matrix

Variable	Conceptual Definition	Indicators	Instrument / Tool	Measurement Scale
Dependent Variable (DV)	The degree of richness, sophistication, and alignment of productive vocabulary with CEFR descriptors in written academic texts.	• TTR (Type-Token Ratio) • MTLD (Measure of Textual Lexical Diversity) • % AWL (Academic Word List) • % K1 - K2 (1k and 2k high-frequency bands)	Compleat Lexical Tutor (Lextutor VP) computational analyzer	Ratio scale (0 - 1) for analytical indices and Percentage (0 - 100%) for frequency bands.
Lexical Profile and Level				

Note. TTR = Type-Token Ratio. AWL = Academic Word List.

2.3. Methodological note on CEFR B2 measurement (response to reviewers)

It is essential to clarify that the baseline B2 level was not determined through an external, standardized test of prior grammatical or receptive knowledge. Instead, it was established indicatively and descriptively through an automated, corpus-based lexical profiling analysis of free-writing samples. This baseline classification is performed empirically by cross-referencing the density of the Academic Word List (AWL) and sophistication indices against the 'criterial features' validated by computational language frameworks for independent users. The Common European Framework of Reference for Languages (CEFR) assesses language proficiency in general skills, so it doesn't set a rigid word limit for writing. However, official exams (such as Cambridge or the DELE) use recommended length standards for texts and essays according to level: A1 (Beginner): Truly short and basic sentences. Writings of 20 to 50 words for filling out forms or writing greetings. A2 (Basic): Short and simple texts. Between 50 and 100 words for notes, messages, or simple emails. B1 (Intermediate): Connected paragraphs and descriptions of experiences. Between 100 and 150 words are recommended. B2 (Upper Intermediate): Essays, formal opinions, and argumentative articles. Between 150 and 250 words are suggested. C1 (Advanced): Structured and detailed texts on complex topics. 250-350 words are expected. C2 (Master's): Extended reports, critical essays, and professional articles. Over 350 words are required, using advanced vocabulary. Students in the fifth and sixth semesters are at a B1 level and are therefore considered for writing an essay of so many words (Council of Europe, 2001).

2.4. Instruments and analysis software

The primary instrument is a written corpus consisting of academic essays of approximately 100–150 words. To guarantee construct and content validity, both the writing task and its analytic assessment rubric were subjected to a panel of expert judges (two experts in English language pedagogy). Utilizing Aiken's V coefficient, the experts verified that the task appropriately demanded the cognitive depth and formal academic register required to elicit the proposed lexical indicators. Kyle et al. (2022) emphasize that corpus-based approaches allow linguistic features to be examined reliably as they naturally occur.

The analysis was performed using LexTutor (VP) and AntConc to determine the lexical frequency profile (K1, K2, and AWL, and Off-list words). According to Abduh & Rosmaladewi (2017), LexTutor helps reveal whether learners rely excessively on high-frequency words or demonstrate control of academic and low-frequency vocabulary. Kyle et al. (2022) stress that automatic tools reduce researcher bias and provide objective, replicable, and multidimensional measurements.

2.5. Corpus preprocessing

Prior to processing the compositions through LexTutor (VP), the dataset underwent a rigorous corpus preprocessing and cleaning stage. This phase focused on the orthographic normalization of lexical variants and phonetically driven fossilized errors identified during a baseline Key Word in Context (KWIC) concordance analysis. Prominent examples detected across student productions included variants such as 'writting' for writing, 'becouse' for because, and 'conclucion' for conclusion. Manual normalization is a mandatory methodological prerequisite in computational corpus linguistics; otherwise, algorithmic parsers misclassify orthographic errors as advanced Off-list words, yielding false positives in lexical sophistication and compromising data reliability.

2.6. Ethical considerations

This research was conducted under strict bioethical guidelines for the processing of linguistic data from human subjects. The data collection and coding process, as well as the informed consent form signed individually by the students, were rigorously reviewed to ensure that the corpus was used for strictly academic, non-punitive purposes, with absolute safeguards of anonymity through alphanumeric coding.

3. Results

The purpose of this investigation is to execute a rigorous linguistic analysis of student-generated academic compositions centered on the thematic axis of "Academic Writing." The analyzed corpus comprises productions from two distinct student cohorts within the language pedagogy framework: 14 essays from the 5th Semester (Parallel "B") and 29 essays from the 6th Semester (Parallel "A").

The corpus linguistics analysis focused on three primary analytical goals:

- Lexical diversity and density: measuring the Type-Token Ratio (TTR) and evaluating the proportion of lexical content words against functional words.
- Vocabulary frequency profiling: identifying occurrences of high-frequency words versus specialized terms from the Academic Word List (AWL).
- CEFR Standards Benchmarking: contrasting metrics against the official proficiency levels (B2/C1) expected by the CEFR in higher education.

3.1. Lexical Diversity (TTR) and density metrics

Table 2

Lexical Diversity (TTR) and Density Performance Summary

Semester / Parallel	Lexical Diversity (Average TTR)	Lexical (%)	Density	High-Frequency Vocabulary (K1– K2)	Academic Word List (AWL) / Technical
5th Semester "B"	0.48	42.1%		88.5%	11.5%
6th Semester "A"	0.56	46.8%		81.2%	18.8%
CEFR Expected Standard (B2/C1)	> 0.60	> 50.0%		< 75.0%	> 25.0%

- Lexical diversity (TTR): compositions in the 5th semester exhibit a TTR of 0.48, signaling a high rate of word repetition. In the 6th semester, the ratio increases to 0.56, demonstrating a wider, more varied selection of vocabulary, although still falling slightly short of the optimal threshold for a fluent user (> 0.60).
- Lexical density: the progression from 42.1% to 46.8% proves that the output of the 6th semester carries a heavier informational load, reducing the over-reliance on basic functional words.

3.2. Vocabulary frequency distribution and CEFR alignment

The corpus analysis software mapped the vocabulary across the CEFR proficiency bands as follows:

- 5th Semester "B": 72% [A1–A2] | 18% [B1] | 10% [B2/AWL]
- 6th Semester "A": 55% [A1–A2] | 24% [B1] | 21% [B2–C1/AWL]

To determine whether the observed variance between the lexical diversity means of the fifth semester (= 0.48) and the sixth semester (= 0.56) carried statistical significance, an inferential independent samples t-test was performed. The parametric contrast yielded an effect size of $d = 0.72$ (Cohen's d), which represents a medium-to-high effect size. This statistical indicator confirms that the lexical progression holds solid significance and demonstrates a measurable, practical impact on student vocabulary maturation between the two consecutive semesters.

Comparative analysis with previous empirical studies when contrasting the present empirical findings with previous literature in similar EFL university contexts, the results align closely with regional and international patterns of academic vocabulary acquisition. The heavy concentration of high-frequency words observed in the UTC corpus (88.5% in 5th semester and 81.2% in 6th semester) mirrors the findings of Tovar (2017) at ESPOL

(Ecuador), who reported that Ecuadorian university learners tend to anchor their written production predominantly in the first 2,000 most frequent word families (K1–K2), struggling to achieve productive control over low-frequency and specialized academic vocabulary.

Furthermore, the average Type-Token Ratio (TTR) values obtained in this study (0.48 for 5th semester and 0.56 for 6th semester) corroborate the cross-sectional lexical trajectory identified by Laufer (2013) and Dang & Webb (2014) who demonstrated that while vocabulary variation increases progressively with higher instructional levels, productive academic fluency remains constrained below the optimal benchmark (> 0.60) unless explicit, targeted lexical instruction is introduced. Regarding the incorporation of the Academic Word List (AWL), the growth from 11.5% to 18.8% between semesters demonstrates clear lexical maturation yet highlights the persistent "lexical plateau" widely documented across Latin American higher education.

This phenomenon coincides with the empirical evidence presented by Hyland & Tse (2007) and Crossley (2020) who emphasized that L2 novice academic writers frequently struggle to surpass an 18–20% threshold of academic lexical items without corpus-informed vocabulary interventions. While the UTC 6th-semester cohort shows substantial progress in utilizing discipline-specific terminology (e.g., *topic sentence*, *coherence*, *comparative*), their compositions still fall short of the autonomous B2/C1 academic writing standard ($> 25\%$ AWL) described by Nation (2018) and Uchida & Negishi (2025). This comparative convergence confirms that the primary constraint in advanced EFL essay writing is not grammatical macrostructure, but the productive mastery of low-frequency academic lemmas and formulaic chunks.

4. Discussion

The primary objective of this investigation was to determine the lexical profile of academic essays produced by pre-service EFL teachers across two curricular cohorts (5th and 6th semesters) at the Technical University of Cotopaxi. Overall, the empirical findings reveal a statistically significant developmental maturation between cohorts—evidenced by moderate-to-large gains in lexical diversity ($d = 0.72$), increased lexical density, and a notable expansion in Academic Word List (AWL) usage. Nevertheless, this observed lexical progression coexists with an enduring reliance on high-frequency vocabulary (K1–K2 bands) and fossilized mechanical errors, highlighting a persistent lexical plateau that prevents most learners from consolidating the independent B2 proficiency demanded by curricular standards. The following discussion interprets these developmental trends in relation to curricular progression, contrasts the observed deficits against the CEFR benchmark, and explores their broader implications for academic writing pedagogy and teacher education.

4.1. Lexical maturation and AWL integration

Quantitative evidence supports a positive lexical evolution across instructional levels. The growth of AWL inclusion from 11.5% to 18.8% in the 6th semester indicates that higher-level students are successfully internalizing academic input and replicating it in free-writing tasks. Rather than relying on purely intuitive definitions, 6th-semester cohorts actively deploy specialized nomenclature such as *topic sentence*, *body paragraphs*, *comparative*, and *cause-effect*. Nevertheless, the finding that over half of the 6th-semester corpus (55%) remains tethered to high-frequency A1–A2 vocabulary points to a persistent reliance on simplified linguistic structures when handling abstract prompts. Cross-referencing these empirical findings with the curriculum design of the PINE major at UTC reveals a direct correlation with course progression. According to the current syllabus, students explicitly take the Academic Writing course during their fifth semester, which explains why the initial corpus begins to display the foundational structural features of essay organization. In the sixth semester, students advance toward courses requiring higher discursive and research complexity. The observed empirical growth of AWL usage represents a net increase of 7.3%, which organically matches the developmental trajectory projected by the department's curricular sequence. However, the average profile still falls short of the optimal 25% target required internationally for autonomous B2 discursive competence.

4.2. Diagnostics of the CEFR Proficiency Gap

University parameters mandate that students in these intermediate-to-advanced stages produce texts firmly anchored within the B2 or C1 bands. Cross-referencing the corpus data with these institutional goals exposes a critical instructional gap:

- Syntactic-lexical deficit: while advanced students frequently grasp the macrostructural elements of an essay (explicitly describing the 5-paragraph model and the utility of an introductory hook), they often lack the lexical density needed to sustain complex arguments with precision.
- Fossilization as a linguistic anchor: the high frequency of flawed word mechanics ("*becouse*", "*knoledge*", "*essencial*") distorts automatic lexical profiling. For computational parsers, the presence of these systematic errors, combined with literal translations from Spanish (e.g., "*with major security*" instead of *greater confidence*), adjusts the actual free-writing competence down to a true B1 level, trailing behind nominal curriculum expectations.
- Outliers within the corpus: outstanding productions such as those by M. Q. (5th "B") and D. L. (6th "A"), defied the general classroom averages, recording over 25% of terms within the B2/C1 range and successfully integrating AWL components like *objective*, *evaluation*, *analytical*, *coherence*, and *structure*.

4.3. Implications for academic writing pedagogy

This diagnostic indicates that traditional instructional methodologies focused exclusively on external text structure (the components of a paragraph or essay) have hit a ceiling. The true educational bottleneck is not a lack of structural comprehension, but rather a lack of

a differentiated, formal lexical repertoire. When students experience a shortage of technical academic vocabulary in English, they display a tendency to default to simplified everyday concepts or shift their focus toward basic pedagogical contexts (such as discussing classroom games or children's prizes) simply to fulfill the required length of the writing prompt.

Finally, it is imperative to acknowledge three core methodological limitations in the present study:

- a) The data collection design relies on a cross-sectional rather than a longitudinal corpus, which restricts the insights to comparing separate groups at a single point in time instead of tracking the chronological development of the same cohort.
- b) The study analyzed a single textual genre (the argumentative opinion essay). While this genre represents the standard assessment tool for university-level academic writing, this focus precludes the generalizability of these lexical profiles to other genres, such as technical reports or narrative texts.
- c) First language transfer (L1 transfer) and literal translation patterns from Spanish to English were not statistically controlled or isolated, representing a cognitive and cross-linguistic variable that may have directly influenced the students' vocabulary choices and structural sequencing.

5. Conclusions

Regarding lexical diversity and density: the findings demonstrate a positive, progressive evolution in students' linguistic maturation as they advance through the curriculum. Sixth-semester students achieve a higher informational load and greater vocabulary variation in their academic essays compared to those in the fifth semester. However, the overall averages reveal that written production still exhibits a considerable degree of lexical redundancy and repetition. This indicates that most students are in a transitional phase and have not yet reached the optimal thresholds of textual fluidity and independence required for advanced academic writing.

Regarding vocabulary frequency profiling and AWL integration: the quantitative corpus analysis reveals a persistent reliance on high-frequency, general English vocabulary, which constitutes more than half of the lexical choices in student compositions. Although the sixth-semester cohort shows an encouraging increase in the appropriation and integration of technical terms and the Academic Word List (AWL), the continued over-reliance on basic vocabulary bands restricts the depth and precision needed to convey complex, discipline-specific concepts, thereby limiting rigorous argumentative development.

Regarding Benchmarking with CEFR Standards: Cross-referencing the corpus data exposes a critical instructional gap between institutional curricular expectations and the actual written production competence of EFL students at UTC. While graduation profiles mandate a proficiency firmly anchored within the independent user bands (B2/C1), the

empirical lexical profile positions the classroom average at a lower intermediate level. This discrepancy is further exacerbated by phonetically driven fossilized errors and orthographic variants, which distort automated lexical measurements and lower overall proficiency ratings.

Regarding Implications for academic writing pedagogy: the diagnostic results empirically prove that traditional instructional methodologies focused exclusively on the external macrostructure of the essay (e.g., the five-paragraph model) have hit a pedagogical ceiling. The primary bottleneck for these future English educators is not a lack of structural comprehension, but rather a deficit in a differentiated, formal academic repertoire. This highlights an urgent need to adjust syllabi and teaching strategies toward the explicit, systematic instruction of academic vocabulary, lexical chunks, and discipline-specific collocations.

6. Conflict of interest

The authors declare that there are no conflicts of interest regarding the submitted article.

7. Copyright

The authors are the holders of the copyright (economic and/or exploitation rights) regarding the journal's content.

8. Authors' Contribution Statement

The authors declare that they have participated equally and substantially in the conduct of this research, under the responsibilities under the [CRediT Taxonomy](#) to describe the individual contributions of each author to the work.

Jessica Dayanna Granja Rubio, carried out the conceptualization of the research, organization and execution of the trial, collection and curation of the data, review of the statistical analysis, interpretation of the results, preparation of the original draft, preparation of tables and figures, critical review of the content and approval of the final version of the manuscript.

Olga Lorena González Ortiz, carried out the academic and methodological supervision of the research, provided guidance during the development of the essay, reviewed the analysis and interpretation of the results, carried out the critical review of the manuscript, made the scientific and academic correction and approved the final version for publication.

Disclosure of delegation to generative AI

The authors declare that they used ChatGPT (OpenAI, GPT-5.5) as a support tool to reorganize and editorially reformulate the original manuscript, correct the style, translate the abstract into English, and check formal compliance with the journal's format. The experimental data, design, statistical results, and scientific decisions come from the

authors' research and were not generated or modified by the tool. The authors reviewed, contrasted, and approved each section and assumed full responsibility for the final content. The tool is not listed as an author and assumes no responsibility for the results. Declaration submitted by: Jessica Dayanna Granja Rubio

9. Funding

This research was funded entirely by the authors' own resources.

10. References

- Abduh, A., & Rosmaladewi, R. (2017) Taking the Lextutor on-line tool to examine students' vocabulary level in business English students. *World Transactions on Engineering and Technology Education*, 15(3), 283-286.
[https://www.wiete.com.au/journals/WTE&TE/Pages/Vol.15,%20No.3%20\(2017\)/15-Abduh-A.pdf](https://www.wiete.com.au/journals/WTE&TE/Pages/Vol.15,%20No.3%20(2017)/15-Abduh-A.pdf)
- Carvajal Lita, M. D. J., Anaguano Pérez, J. C., & Garzón Puetate, W. A. (2025). Tendencias y conexiones temáticas entre economía social, innovación social, emprendimiento social y desarrollo sostenible: un análisis bibliométrico (2009–2024). *Conectividad*, 6(4), 197–215.
<https://doi.org/10.37431/conectividad.v6i4.409>
- Conti, G. (2025). *Why teaching chunks in MFL is more effective for developing fluency: a research-based perspective*. The Language Gym.
<https://gianfrancoconti.com/2025/03/11/why-teaching-chunks-in-mfl-is-more-effective-for-developing-fluency-a-research-based-perspective/>
- Council of Europe. (2001). *Common European framework of reference for languages: learning, teaching, assessment*. Council of Europe Publishing, Strasbourg.
<https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches (5th edition)*. Sage publications.
[https://books.google.com.ec/books?id=335ZDwAAQBAJ&lpg=PT16&ots=YEA YMksssN&dq=Creswell%2C%20J.%20W.%2C%20%26%20Creswell%2C%20J.%20D.%20\(2017\).%20Research%20design%3A%20Qualitative%2C%20quantitative%2C%20and%20mixed%20methods%20approaches.%20Sage%20publications&lr&pg=PT16#v=onepage&q&f=false](https://books.google.com.ec/books?id=335ZDwAAQBAJ&lpg=PT16&ots=YEA YMksssN&dq=Creswell%2C%20J.%20W.%2C%20%26%20Creswell%2C%20J.%20D.%20(2017).%20Research%20design%3A%20Qualitative%2C%20quantitative%2C%20and%20mixed%20methods%20approaches.%20Sage%20publications&lr&pg=PT16#v=onepage&q&f=false)
- Crossley, S. A. (2020). Linguistic features in writing quality and development: an overview. *Journal of Writing Research*, 11(3), 415-443.
<https://www.jowr.org/index.php/jowr/article/view/582>
- Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, 33, 66-76. <https://doi.org/10.1016/j.esp.2013.08.001>

- Etaat, F. (2026). Exploring linguistic fingerprints in human and AI-generated texts: an nlp-based approach in second language writing. *Ampersand (Oxford, UK)*. 16(100258). 100258. <https://doi.org/10.1016/j.amper.2026.100258>
- Fredrick, D. R. & Craven G. (2025). Lexical diversity, syntactic complexity, and readability: a corpus-based analysis of L2 student essays. *Frontiers in Education*, 10(1616935). <https://doi.org/10.3389/educ.2025.1616935>
- Hyland, K., & Tse, P. (2007). ¿Existe un “vocabulario académico”? *TESOL Quarterly*, 41, 235-253. <https://onlinelibrary.wiley.com/doi/10.1002/j.1545-7249.2007.tb00058.x>
- Hunziker, S., & Blankenagel, M. (2021). Research design in business and management: a practical guide to cross-sectional and longitudinal research. *Springer Nature*. <https://doi.org/10.1007/978-3-658-34357-6>
- Kyle, K., Crossley, S. A., & Berger, C. M. (2022). The Tool for the Automatic Analysis of Lexical Sophistication (TAALES): Version 2.2. *Behavior Research Methods*, 54(1), 1030–1046. <https://doi.org/10.3758/s13428-017-0924-4>
- Laufer, B. (2013) Vocabulary and Writing. In: *The Encyclopedia of Applied Linguistics*. Wiley. <https://doi.org/10.1002/9781405198431.wbeal1432>
- Lu, X., & Hu., R. (2022). Sense-aware lexical sophistication indices and their relationship to second language writing quality. *Behavior Research Methods* 54(3), 1444–1460. <https://doi.org/10.3758/s13428-021-01675-6>
- Lyashevskaya, O. N., Pyzhak, J. V., & Vinogradova, O. I. (2022). Word-formation complexity: a learner corpus-based study. *Russian Journal of Linguistics*, 26(2), 471-492. <https://doi.org/10.22363/2687-0088-31187>
- Nation, I. S. P. (2018). *Learning Vocabulary in Another Language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
- Park, Y. S., Konge, L. & Artino, A. R. Jr. (2020). The positivism paradigm of research. *Academic Medicine: Journal of the Association of American Medical Colleges*, 95(5), 690-694. <https://doi.org/10.1097/ACM.0000000000003093>
- Read, J. (2010). *Assessing Vocabulary*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511732942>
- Uchida, S., & Negishi, M. (2025). Assigning CEFR-J levels to English learners' writing: an approach using lexical metrics and generative AI. *Research Methods in Applied Linguistics*, 4(2), 100199. <https://kyushu-u.elsevierpure.com/en/publications/assigning-cefr-j-levels-to-english-learners-writing-an-approach-u/>
- Schmitt, N. (2010). *Researching vocabulary: a vocabulary research manual*. Palgrave Macmillan. <http://dx.doi.org/10.1057/9780230293977>

Siedlecki, S. L. (2020). Understanding descriptive research designs and methods. *Clinical Nurse Specialist*, 34(1), 8-12.

<https://doi.org/10.1097/NUR.0000000000000493>

Tovar Viera, R. (2017). Vocabulary knowledge in the production of written texts: a case study on EFL language learners. *Revista Tecnológica – ESPOL*, 30(3).

<https://www.rte.espol.edu.ec/index.php/tecnologica/es/article/view/628>

El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Explorador Digital**.



El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Explorador Digital**.

