# Exploring lexical features in IELTS Writing Task 2 essays: A case of Vietnamese non-English-major students

***Trieu Vy Truong\****

*\*Saigon International University*

**Abstract:** *This quantitative study examines the lexical features of IELTS Academic Writing Task 2 (IAWT2) essays written by Vietnamese non-English-major undergraduates, focusing on following writing prompts (WP): WP1 (encouraging bicycle use), WP2 (urban living and health), and WP3 (preferences for home-cooked versus fast food). A corpus of 479 essays was analyzed to explore the relationships between lexical density and content word types. Results reveal that nouns and adjectives significantly enhance lexical density, while verbs and adverbs contribute minimally. These results underscore the need for task-specific lexical training to improve students' academic writing proficiency and vocabulary use in IAWT2.*

**Keywords:** *IELTS Writing Task 2; lexical features; learner corpora; Vietnamese non-English-major students;*

## 1. Introduction

The IELTS Academic Writing Task 2 (IAWT2) assesses candidates' ability to write structured essays on complex topics, requiring precise use of a wide range of vocabulary, including uncommon and academic words (Cullen et al., 2014; Nguyen, 2016; IELTS, 2023). These demands present challenges for EFL learners, who struggle with lexical density, diversity, and sophistication (Ariani & Arham, 2020; Ashraf et al., 2024). Learner corpora and tools like AntConc and TAALED offer valuable insights into these lexical features, aiding students in improving their writing (Pérez-Paredes & Mark, 2022; Kyle et al., 2020).

Studies show that lexical sophistication correlates with better academic writing, though challenges like reliance on high-frequency words persist (Vyatkina & Housen, 2020). In Vietnam, lexical diversity and sophistication influence essay quality, though topic familiarity also plays a role (Nguyen et al., 2022; Pham & Tran, 2023). Research gaps include the relationship between content word types and lexical density, crucial for assessing text complexity. This study addresses these gaps by analyzing the relationship between lexical density and content words in each writing prompt in 479 IAWT2 essays by Vietnamese non-English-major students, aiming to enhance their writing performance.

## 2. Findings and discussion
### 2.1. Findings

*Table 1: Descriptive statistics of lexical density and distribution of content word types*

| Writing prompt | Lexical density | | Distribution of content word types (ratio) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | noun | | verb | | adjective | | adverbs | |
| | M | SD | M | SD | M | SD | M | SD | M | SD |
| WP1 | .55 | .05 | .28 | .03 | .16 | .02 | .09 | .03 | .05 | .01 |
| WP2 | .58 | .04 | .29 | .02 | .13 | .02 | .12 | .02 | .04 | .01 |
| WP3 | .56 | .04 | .24 | .03 | .15 | .02 | .11 | .02 | .05 | .02 |

Lexical density was slightly higher in WP2 (M = .58, SD = .04) compared to WP1 (M = .55, SD = .05) and WP3 (M = .56, SD = .04), indicating consistency across tasks. Nouns dominated content word types in all prompts, with WP2 (M = .29, SD = .02) slightly higher than WP1 (M = .28, SD = .03) and WP3 (M = .24, SD = .03). The lower noun proportion in WP3 might reflect a shift toward action or description.

Verbs were most frequent in WP1 (M = .16, SD = .02), followed by WP3 (M = .15, SD = .02) and WP2 (M = .13, SD = .02), suggesting a more dynamic writing style in WP1. Adjectives were slightly higher in WP2 (M = .12, SD = .02) compared to WP3 (M = .11, SD = .02) and WP1 (M = .09, SD = .03), reflecting more descriptive language in WP2. Adverbs were used sparingly across all prompts, with WP1 and WP3 (M = .05) slightly higher than WP2 (M = .04), showing minimal variation by prompt.

*Table 2: Oneway-ANOVA between LD and content word types in WP1*

| | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| noun | Between Groups | .147 | 145 | .001 | 3.084 | .013 |
| | Within Groups | .004 | 13 | .000 | | |
| | Total | .151 | 158 | | | |
| verb | Between Groups | .069 | 145 | .000 | 1.509 | .204 |
| | Within Groups | .004 | 13 | .000 | | |
| | Total | .073 | 158 | | | |
| adj | Between Groups | .106 | 145 | .001 | 9.155 | .000 |
| | Within Groups | .001 | 13 | .000 | | |
| | Total | .107 | 158 | | | |
| adv | Between Groups | .045 | 145 | .000 | 1.573 | .179 |
| | Within Groups | .003 | 13 | .000 | | |
| | Total | .048 | 158 | | | |

*Table 3: Correlation between LD and content word types in WP1*

| | | LD | noun | verb | adj | adv |
|---|---|---|---|---|---|---|
| LD | Pearson Correlation | 1 | .669** | .178* | .563** | -.194* |
| | Sig. (2-tailed) | | .000 | .025 | .000 | .015 |
| | N | 159 | 159 | 159 | 159 | 159 |

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

The ANOVA results (Table 2.) revealed significant differences in the use of nouns, $F_{(145, 13)} = 3.08$, p = .013, and adjectives, $F_{(145, 13)} = 9.16$, p < .001, indicating variability in lexical density for these word types in argumentative essays. In contrast, verbs, $F_{(145, 13)} = 1.51$, p = .204, and adverbs, $F_{(145, 13)} = 1.57$, p = .179, showed no significant differences, suggesting consistency in their usage in WP1.

Pearson correlation analysis (Table 3.) further supported these findings. LD was strongly correlated with nouns (r = .669, p < .001) and moderately with adjectives (r = .563, p < .001), reinforcing their key role in lexical variation. Verbs exhibited a weak positive correlation with LD (r = .178, p = .025), while adverbs showed a weak negative correlation (r = -.194, p = .015), aligning with their more stable patterns observed in the ANOVA.

*Table 4: ANOVA between LD and content word types in WP2*

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| noun | Between Groups | .126 | 142 | .001 | 7.554 | .000 |
| | Within Groups | .002 | 17 | .000 | | |
| | Total | .128 | 159 | | | |

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| verb | Between Groups | .047 | 142 | .000 | 6.298 | .000 |
| | Within Groups | .001 | 17 | .000 | | |
| | Total | .048 | 159 | | | |
| adj | Between Groups | .067 | 142 | .000 | 79.875 | .000 |
| | Within Groups | .000 | 17 | .000 | | |
| | Total | .067 | 159 | | | |
| adv | Between Groups | .029 | 142 | .000 | 5.403 | .000 |
| | Within Groups | .001 | 17 | .000 | | |
| | Total | .030 | 159 | | | |

*Table 5: Correlation between LD and content word types in WP2*

| | | LD | noun | verb | adj | adv |
|---|---|---|---|---|---|---|
| LD | Pearson Correlation | 1 | .666** | -.302** | .537** | -.268** |
| | Sig. (2-tailed) | | .000 | .000 | .000 | .001 |
| | N | 160 | 160 | 160 | 160 | 160 |

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Table 4. demonstrates statistically significant differences in the usage of all content word types across groups: nouns, $F_{(142, 17)} = 7.55$, p < .001; verbs, $F_{(142, 17)} = 6.30$, p < .001; adjectives, $F_{(142, 17)} = 79.88$, p < .001; and adverbs, $F_{(142, 17)} = 5.40$, p < .001. These results underscore considerable variability in lexical density and content word distribution among groups in this writing prompt.

This variability is further clarified by the correlation analysis in WP2 (Table 5.), which revealed strong positive associations between LD and nouns (r = .666, p < .001), as well as moderate positive correlations with adjectives (r = .537, p < .001). Conversely, LD was moderately negatively correlated with verbs (r = -.302, p < .001) and adverbs (r = -.268, p = .001). Together, these findings suggest that nouns and adjectives drive higher lexical density, while verbs and adverbs are less prominent in lexically dense writing.

*Table 6: ANOVA between LD and content word types in WP3*

| | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| noun | Between Groups | .146 | 152 | .001 | 33.515 | .000 |
| | Within Groups | .000 | 7 | .000 | | |
| | Total | .146 | 159 | | | |
| verb | Between Groups | .067 | 152 | .000 | . | . |
| | Within Groups | .000 | 7 | .000 | | |
| | Total | .067 | 159 | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| adj | Between Groups | .070 | 152 | .000 | . | . |
| | Within Groups | .000 | 7 | .000 | | |
| | Total | .070 | 159 | | | |
| adv | Between Groups | .038 | 152 | .000 | 2.207 | .133 |
| | Within Groups | .001 | 7 | .000 | | |
| | Total | .039 | 159 | | | |

*Table 7: Correlation between LD and content word types in WP3*

| | | LD | noun | verb | adj | adv |
|---|---|---|---|---|---|---|
| LD | Pearson Correlation | 1 | .727** | -.421** | .461** | -.264** |
| | Sig. (2-tailed) | | .000 | .000 | .000 | .001 |
| | N | 160 | 160 | 160 | 160 | 160 |
| **\*\*. Correlation is significant at the 0.01 level (2-tailed).** <br> **\*. Correlation is significant at the 0.05 level (2-tailed).** | | | | | | |

Table 6. highlights a significant difference in noun usage across groups, $F_{(152, 7)} = 33.52$, $p < .001$, while adverb usage remains consistent, $F_{(152, 7)} = 2.21$, $p = .133$. These findings align with the correlation analysis in WP3 , which shows a strong positive relationship between LD and nouns ($r = .727$, $p < .001$), and a moderate positive correlation with adjectives ($r = .461$, $p < .001$). Conversely, LD is moderately negatively correlated with verbs ($r = -.421$, $p < .001$) and adverbs ($r = -.264$, $p = .001$), reinforcing the role of nouns and adjectives in driving lexical density, while verbs and adverbs diminish in prominence.

Overall, these results from WP3 align with patterns observed in WP1 and WP2, demonstrating consistent roles of nouns and adjectives as key contributors to lexical density (LD) across writing prompts. In WP1, a strong positive correlation between LD and nouns ($r = .669$) and a moderate positive correlation with adjectives ($r = .563$) were reported, similar to the findings in WP3 ($r = .727$ and $r = .461$, respectively). WP2 further reinforced this trend, showing comparable correlations (nouns: $r = .666$, adjectives: $r = .537$).

The moderate negative correlations between LD and verbs (WP1: $r = .178$; WP2: $r = -.302$; WP3: $r = -.421$) and adverbs (WP1: $r = -.194$; WP2: $r = -.268$; WP3: $r = -.264$) also reflect a consistent reduction in these word types' usage as LD increases. Collectively, these results suggest that regardless of the writing prompt, nouns and adjectives are pivotal in enhancing lexical density, while verbs and adverbs play a diminishing role in more lexically dense texts.

## 2.2. Discussion

This study examines the lexical features in IAWT2 essays written by Vietnamese non-English-major students, highlighting significant correlations between lexical density and nouns across all writing prompts (e.g., $r = .727$ in WP3). These findings confirm earlier studies associating a high proportion of nouns with informational density in academic writing (Kopple, 1994; Vidakovic & Barker, 2010). By focusing on non-English-major students, the study contributes to understanding how this demographic constructs informationally rich texts, with nuanced relationships such as the moderate contribution of adjectives and limited roles of verbs and adverbs adding depth to existing knowledge.

This study has some limitations. First, the one-week interval between writing tasks may not have allowed participants sufficient time to develop their lexical skills. Future research could adopt a longitudinal learner corpus approach to track vocabulary changes over extended periods (Szudarski, 2018; Timmis, 2015). Second, this study relied solely on quantitative analysis, which, while insightful, may overlook some nuanced aspects of lexical features. Future research should incorporate qualitative methods, such as analyzing the collocation patterns in students essay to see how content words types interact with each other.

## 3. Conclusion

This study provides valuable insights into the lexical features of IAWT2 essays written by Vietnamese non-English-major undergraduates. Analyzing a corpus of 479 essays, the research highlights significant relationships between lexical density and content word types. Nouns and adjectives emerged as pivotal in enhancing lexical density, while verbs and adverbs contributed less prominently. These findings emphasize the critical role of task-specific vocabulary training in preparing students for IAWT2. Future research could extend these findings through longitudinal studies and incorporate qualitative analyses of collocation patterns to deepen understanding of lexical development.

## References

1. Kopple, W. J. V. (1994). Some characteristics and functions of grammatical subjects in scientific discourse. Written Communication, 11(4), 534–564. https://doi.org/10.1177/0741088394011004004

2. Szudarski, P. (2018). Corpus linguistics for vocabulary: A guide for research. Routledge.

3. Timmis, I. (2015). Corpus linguistics for ELT: Research and practice. Routledge.

4. Vidakovic, I., & Barker, F. (2010). Use of words and multi-words unites in Skills for Life in writing examination. In IELTS Research Report 41. IELTS Australian/British Council.