
A corpus-based analysis of lexical diversity (TTR) in fourth-year Uzbek EFL students' essays

Xayrullayeva Charosxon Sayfullo kizi
khayrullaevacharos2@gmail.com
Student of 3rd English faculty,
Uzbekistan State World Languages University

Scientific advisor: **Gulomova Raykhona Rustamjonovna**
rayhonagulomova@gmail.com
Senior teacher,
Uzbekistan State World Languages University

Annotation *Vocabulary knowledge plays a crucial role in developing writing proficiency in a second language. Among various lexical measures, lexical diversity is widely used to evaluate learners' vocabulary range in written production. This study examines lexical diversity in IELTS Task 2 essays written by fourth-year Uzbek EFL students using a corpus-based approach. A small learner corpus consisting of 10 essays, each containing a minimum of 250 words, was compiled and analyzed. Lexical diversity was measured using the Type-Token Ratio (TTR). The corpus was processed using AntConc software to determine the total number of word tokens and unique word types. The analysis revealed a moderate level of lexical diversity with a TTR score of 0.29. This result indicates that students are able to use a relatively varied vocabulary in their academic writing, although the repetition of high-frequency lexical items remains noticeable. The findings contribute to a better understanding of vocabulary use in Uzbek EFL academic writing and highlight the importance of continued vocabulary development even at advanced stages of language learning.*

Keywords *Lexical diversity, TTR, corpus-based study, EFL writing, Uzbek learners*

To'rtinchi bosqich o'zbek EFL talabalari insholarida leksik xilma-xillik (TTR)ning korpus asosida tahlili

Xayrullayeva Charosxon Sayfullo qizi
khayrullaevacharos2@gmail.com
Ingliz tili 3-fakulteti,
O'zbekiston davlat jahon tillari universiteti

Ilmiy rahbar: **G'ulomova Raykhona Rustamjonova**
gulomovarayhona@gmail.com
Katta o'qituvchi,
O'zbekiston davlat jahon tillari universiteti

Annotatsiya *Lug'at boyligi ikkinchi tilni o'rganishda, ayniqsa yozma nutq ko'nikmalarini rivojlantirishda muhim rol o'ynaydi. Turli leksik ko'rsatkichlar orasida leksik xilma-xillik (lexical diversity) o'quvchilarning yozma matnlarda ishlatadigan lug'at boyligini baholashda keng qo'llaniladigan muhim mezonlardan biridir. Ushbu tadqiqot to'rtinchi kurs o'zbek EFL talabalari tomonidan yozilgan IELTS Task 2 insholarida leksik xilma-xillik darajasini korpusga asoslangan yondashuv yordamida tahlil qilishga qaratilgan. Tadqiqot uchun kamida 250 so'zdan iborat bo'lgan 10 ta*

inshodan tashkil topgan kichik o'quvchi korpusi tuzildi va tahlil qilindi. Leksik xilma-xillik Type-Token Ratio (TTR) ko'rsatkichi orqali o'lchandi. Korpus AntConc dasturi yordamida qayta ishlanib, umumiy so'zlar soni (tokenlar) hamda noyob so'zlar soni (typelar) aniqlandi. Tahlil natijalari leksik xilma-xillikning o'rtacha darajada ekanligini (TTR = 0.29) ko'rsatdi. Bu natija talabalar akademik yozuvlarda ma'lum darajada turli xil lug'atdan foydalanishini ko'rsatadi, biroq yuqori chastotali so'zlarning takrorlanishi hali ham kuzatiladi. Tadqiqot natijalari o'zbek EFL talabalari akademik yozuvlarida lug'at qo'llanilishi haqida muhim ma'lumotlar beradi hamda yuqori bosqichlarda ham lug'at boyligini rivojlantirish zarurligini ta'kidlaydi.

Kalit so'zlar *Leksik xilma-xillik, TTR, korpusga asoslangan tadqiqot, EFL yozuvi, o'zbek talabalari*

Корпусный анализ лексического разнообразия (TTR) в эссе узбекских студентов четвёртого курса, изучающих английский язык как иностранный (EFL)

Хайруллаева Чаросхон Сайфулло кизи
khayrullaevacharos2@gmail.com

*Студент 3-го факультета английского языка,
Узбекский государственный университет
мировых языков*

Научный руководитель: **Гуломова Райхона Рустамжоновна**
rayhonagulomova@gmail.com
*Старший преподаватель,
Узбекский государственный университет
мировых языков*

Аннотация

Словарный запас играет ключевую роль в формировании письменной компетенции на втором языке. Среди различных лексических показателей лексическое разнообразие (lexical diversity) широко используется для оценки объёма словарного запаса учащих в письменной речи. В данном исследовании анализируется лексическое разнообразие в эссе IELTS Task 2, написанных студентами четвёртого курса из Узбекистана, изучающими английский язык как иностранный (EFL), с использованием корпусного подхода. Для анализа был создан небольшой учебный корпус, состоящий из 18 эссе объёмом не менее 250 слов каждое. Лексическое разнообразие измерялось с помощью коэффициента Type-Token Ratio (TTR). Корпус был обработан с использованием программы AntConc для определения общего количества слов (tokens) и количества уникальных слов (types). Результаты показали средний уровень лексического разнообразия (TTR = 0.27), что свидетельствует о владении студентами функциональной академической лексикой, однако повторение высокочастотных слов остаётся заметным. Полученные данные способствуют лучшему пониманию использования словарного запаса в академическом письме узбекских студентов EFL и подчёркивают необходимость дальнейшего развития лексических навыков на продвинутом этапе обучения.

Ключевые слова *Лексическое разнообразие, TTR, корпусное исследование, письменная речь EFL, узбекские студенты*

Introduction

Vocabulary knowledge is widely known as a fundamental component of second language (L2) proficiency. A broader vocabulary enables learners to express ideas more precisely and construct more coherent and sophisticated texts (Nation, 2001). In academic context, lexical richness is often accepted as a crucial indicator in evaluating writing proficiency.

In the field of English as a Foreign Language (EFL), researchers frequently evaluate learners' writing by examining lexical features such as diversity, sophistication, and density (Laufer & Nation, 1995). However, among all these features, lexical diversity has received more attention comparing to others, because it reflects the range of vocabulary used in a text.

In recent years, English language education in Uzbekistan has experienced considerable reform and growth. Nevertheless, there is still a lack of academic studies focused on lexical richness of Uzbek EFL learners' writing. Investigating lexical diversity among senior year undergraduate students may offer important insights into their vocabulary development and overall writing proficiency.

The main purpose of this study is examining lexical diversity in academic essays written by fourth-year Uzbek EFL students by using a corpus-based approach.

The research question which guiding this study is:

What is the level of lexical diversity (measured by TTR) in fourth-year Uzbek EFL students' academic essays?

Literature review

Lexical richness is a multi-component construct covering 3 main groups: lexical diversity, lexical sophistication, and lexical density (Nation, 2001). Lexical diversity means

the range and variability of vocabulary that is used in a text.

One of the most commonly used measures of lexical diversity is the Type-Token Ratio (TTR). TTR is calculated by dividing the number of word types (unique words) by the number of word tokens (total running words). A higher TTR indicates wider lexical variation.

TTR has certain limitations such as its sensitivity to text length, while it is referred commonly in small-scale corpus studies due to its simplicity and transparency (Malvern et al., 2004). For relatively short texts of similar length, TTR provides a practical and accessible measure of vocabulary range. Therefore, it was selected as the primary analytical tool for this research.

Research has consistently demonstrated a relationship between lexical richness and writing proficiency. It means if learners have larger vocabularies, they tend to produce more varied and sophisticated texts (Laufer & Nation, 1995). Therefore, checking lexical diversity has become a valuable tool in evaluating EFL writing performance (Crossley & McNamara, 2012).

Although the number of international researches is growing, the studies focusing specifically on Uzbek EFL learners is still limited. A corpus-based analysis of lexical diversity in this context can contribute to filling this research gap.

Corpus linguistics provides systematic methods for analyzing authentic language data. By gathering learner texts into a corpus, researchers are able to evaluate objectively patterns of vocabulary use (McEnery & Hardie, 2012). Corpus-based methods develop reliability and validity, as it relies on quantitative data rather than subjective evaluation.

Given these advantages, a corpus-based approach was adopted in the present study.

Methodology

This study used a corpus-based quantitative method to analyze lexical diversity in learner's writing. This research is mainly focused on measuring vocabulary variation using TTR.

The data-set consisted of 10 IELTS Task 2 essays written by fourth-year Uzbek EFL students during regular classroom instruction. Each essay contained a minimum of 250 words and addressed argumentative prompts which is typical for IELTS examinations.

Essay topics varied across students. All identifying information was removed to ensure anonymity. Each essay was converted into plain text (.txt) format before analysis.

The compiled learner corpus contained a total of 2,962 tokens. Using AntConc, the Word List function generated frequency lists and identified 870 word types.

Lexical diversity was calculated using the formula:

$$\text{TTR} = \text{Types} / \text{Tokens}$$

Results

The analysis revealed a total of 2960 tokens and 870 types in the corpus. The overall TTR score was 0.29.

The frequency list indicated repeated use of high-frequency nouns such as "people," "education," "government," and "society." Modal verbs such as "should" and "can" were also frequently observed, reflecting the argumentative nature of IELTS Task 2 essays.

Measure	Value
Total tokens	2960
Total types	870
TTR	0.29

Table 1. *Lexical diversity results of the learner corpus*

Discussion

The Type–Token Ratio (TTR) value of 0.29 indicates a moderate level of lexical diversity in the IELTS Task 2 essays written by fourth-year Uzbek EFL students. This result suggests that although students are capable of producing coherent written texts and using appropriate academic vocabulary, their lexical choices tend to rely on a relatively limited range of frequently used words. In other words, while students demonstrate functional control of academic language, the repetition of certain lexical items reduces the overall variation in vocabulary use.

These findings are consistent with earlier research on second language writing, which suggests that many EFL learners depend on a restricted set of familiar lexical items when composing academic texts. For example, Laufer and Nation (1995) noted that language learners

often repeat common vocabulary items rather than using a wider range of synonyms or less frequent words. In the present study, students were able to construct well-organized argumentative essays, yet the overall vocabulary range did not fully reflect the level of lexical richness typically associated with advanced language proficiency.

Another factor that may have influenced the results is the variation in essay topics. Since IELTS Task 2 prompts require students to respond to different issues, the topic of each essay may guide the selection of certain vocabulary items. Consequently, topic-related lexical choices may have contributed to the repetition of specific words and expressions, which in turn affects the lexical diversity score.

It is also important to note that the Type–Token Ratio is sensitive to corpus size. Because

the dataset used in this study was relatively small, the results should be interpreted with caution. As Malvern et al. (2004) point out, smaller corpora can influence TTR values and may limit the generalizability of findings. Nevertheless, despite this limitation, the analysis still provides valuable insights into vocabulary use among advanced Uzbek EFL learners and highlights important patterns in their academic writing.

Conclusion

This study investigated lexical diversity in IELTS Task 2 essays written by fourth-year Uzbek EFL students using a corpus-based analytical approach. The findings revealed a moderate level of lexical diversity, indicating that while students possess a functional command of academic vocabulary and are able to construct coherent argumentative essays, their vocabulary range still shows certain limitations. The tendency to repeat commonly used lexical items suggests that further development in lexical variation and vocabulary sophistication is needed.

The study contributes to a better understanding of vocabulary use in the Uzbek EFL context and emphasizes the importance of expanding learners' lexical repertoire, particularly at advanced stages of language learning. Developing a wider range of vocabulary may help students express ideas more precisely and produce more lexically rich academic texts.

Future research could build on these findings by analyzing larger and more diverse corpora of student writing. Comparative studies involving learners at different proficiency levels or from different educational contexts could also provide additional insights. Furthermore, incorporating additional lexical measures such as lexical sophistication, lexical density, MTL (Measure of Textual Lexical Diversity), or HD-D could offer a more comprehensive evaluation of vocabulary development in EFL academic writing. Such approaches would help researchers and educators better understand how lexical diversity evolves in advanced language learning contexts.

References:

1. Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and lexical sophistication. *Journal of Research in Reading*, 35(2), 115–135. <https://doi.org/10.1111/j.1467-9817.2010.01449.x>
2. Jarvis, S. (2013). *Defining and measuring lexical diversity*. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley-Blackwell.
3. Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
4. Malvern, D., Richards, B., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan.
5. McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
6. Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.