HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

A CONTRASTIVE ANALYSIS OF LEXICAL DIVERSITY IN HUMAN-WRITTEN AND AI-GENERATED ENGLISH AND UZBEK ADVERTISING TEXTS

Khakimzoda Maftuna
E-mail: khakimzodamaftuna@gmail.com
Doi: https://doi.org/10.5281/zenodo.22970669

Abstract. This study investigates lexical diversity in human-written and AI-generated advertising texts in English and Uzbek from a contrastive linguistic perspective. The research focuses on how artificial intelligence influences vocabulary selection, lexical repetition, semantic variation, and persuasive language in advertising discourse. Particular attention is paid to differences between English and Uzbek advertising texts and to the extent to which AI-generated advertisements reproduce or modify the lexical patterns characteristic of human-authored advertising. The study proposes a corpus-based analytical framework involving Type-Token Ratio (TTR), Moving Average Type-Token Ratio (MATTR), and Measure of Textual Lexical Diversity (MTLD), supplemented by qualitative analysis of lexical fields, evaluative vocabulary, promotional collocations, borrowings, cultural markers, and repetitive patterns. Since TTR is sensitive to text length, the incorporation of length-resistant measures such as MATTR and MTLD provides a more reliable basis for comparison. The analysis suggests that human-written advertisements tend to demonstrate greater contextual and culturally motivated lexical variation, whereas AI-generated texts may display more formulaic patterns and recurrent promotional vocabulary. At the same time, AI-generated texts can demonstrate considerable lexical range, particularly when prompted to produce creative or audience-specific advertising. The English and Uzbek datasets also reveal differences related to cultural values, morphological structure, borrowing, and persuasive strategies. The study contributes to contemporary research on AI-mediated discourse, advertising linguistics, and contrastive lexical analysis.

Keywords: lexical diversity, artificial intelligence, advertising discourse, human-written texts, AI-generated texts, English, Uzbek, contrastive analysis.

INTRODUCTION.

Advertising is a highly strategic form of communication in which linguistic choices are used to attract attention, construct brand identity, create emotional associations, and influence consumer behavior. Unlike ordinary informational discourse, advertising language is characterized by a high degree of communicative economy: advertisers must communicate a substantial amount of meaning within a limited textual space. Consequently, lexical selection becomes an important component of advertising effectiveness. English and Uzbek advertising texts share several universal persuasive mechanisms, including positive evaluation, imperative constructions, emotional vocabulary, metaphorical expressions, repetition, and concise slogans. Nevertheless, their lexical realization is influenced by differences in linguistic structure and sociocultural context. Recent comparative research on English and Uzbek advertising has identified differences in lexical, structural, stylistic, and pragmatic organization, while also emphasizing the importance of culturally specific persuasive strategies. The rapid development of generative artificial intelligence has introduced a new dimension to advertising discourse. Large language models can now produce slogans, product descriptions,

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

social-media advertisements, promotional emails, and other forms of commercial communication within seconds. This development raises important linguistic questions: Do AI-generated advertisements employ lexical diversity comparable to human-written advertisements? Does AI-generated advertising rely more heavily on repeated promotional formulas? Are the differences between AI and human writing the same in English and Uzbek? How does the morphological structure of Uzbek influence lexical diversity in AI-generated texts? These questions are particularly relevant because lexical diversity is not simply a measure of the number of different words occurring in a text. It also reflects how vocabulary is distributed, repeated, and functionally employed. Contemporary research recommends using multiple measures rather than relying exclusively on the traditional Type-Token Ratio (TTR), because TTR is strongly affected by text length. MTLD and other measures provide more stable estimates of lexical diversity across texts of different lengths.

The present study therefore proposes a contrastive framework for analyzing lexical diversity in four categories of advertising discourse: human-written English advertising texts, AI-generated English advertising texts, human-written Uzbek advertising texts, AI-generated Uzbek advertising texts. The principal objective is to identify similarities and differences in lexical diversity, repetition, lexical choice, and persuasive vocabulary across these four categories.

LITERATURE REVIEW

Lexical diversity as a linguistic variable. Lexical diversity refers broadly to the range and distribution of different lexical items within a text. The traditional Type-Token Ratio is calculated by dividing the number of unique lexical items, or types, by the total number of running words, or tokens:

$$TTR = V / N \times 100$$

where V represents the number of unique word types and N represents the total number of tokens.

Although TTR is easy to calculate and interpret, its major limitation is its sensitivity to text length. As a text becomes longer, repeated vocabulary tends to increase, causing TTR to decline even when the underlying lexical repertoire remains comparable. For this reason, researchers have developed alternative measures. MATTR calculates TTR across moving windows and is less dependent on overall text length. MTLD measures the average length of sequential word strings that maintain a specified level of lexical variation. A validation study by McCarthy and Jarvis found that MTLD performed well across several validity dimensions and was less affected by text length than several competing measures. Recent research also suggests that lexical diversity can be examined by grammatical category. For example, noun diversity and verb diversity may reveal patterns that an overall lexical-diversity index cannot capture.

Lexical characteristics of advertising discourse. Advertising language combines informational and persuasive functions. Lexical items are selected not only for their denotative meanings but also for their emotional, evaluative, cultural, and associative potential. English advertising frequently employs concise and memorable lexical constructions, including adjectives such as *new, natural, smart, powerful, fresh, premium,* and *exclusive*. It also makes extensive use of imperatives and action-oriented verbs such as *discover, experience, choose, enjoy,* and *try*. Uzbek advertising similarly employs evaluative and persuasive vocabulary but may demonstrate stronger use of culturally recognizable expressions, honorific forms, emotionally marked words, and locally meaningful references. Recent comparative studies describe English advertising as tending toward brevity and

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

creativity, while Uzbek advertising may incorporate national-cultural symbols and expressions associated with local values.

AI-generated language and lexical diversity. The emergence of large language models has made AI-generated text an increasingly important object of linguistic research. Recent research comparing ChatGPT-generated and human-written English texts found significant differences across several dimensions of lexical diversity, suggesting that AI-generated texts cannot simply be assumed to reproduce human lexical behavior. However, lexical diversity in AI-generated advertising requires a more specific analysis because advertising itself is highly formulaic. Words and expressions such as *best, quality, discover, experience, innovative,* and *perfect* may be repeated intentionally because they fulfill persuasive and branding functions. Therefore, a lower level of lexical diversity should not automatically be interpreted as inferior language production. In advertising, controlled lexical repetition can improve memorability and reinforce a brand message. The linguistic task is consequently to distinguish between functional repetition and excessive formulaicity.

RESULTS AND DISCUSSION

General differences between human-written and AI-generated advertising. The proposed analysis indicates that human-written and AI-generated advertising texts can differ in the distribution of lexical items even when they address the same product category. Human-written advertising is expected to demonstrate greater contextual specificity. Human copywriters can incorporate locally relevant references, unexpected lexical combinations, humor, cultural associations, and product-specific terminology. Such lexical choices may increase variation without necessarily increasing text length.

AI-generated advertising, by contrast, may rely more frequently on established promotional patterns. Examples include combinations such as:

- *discover the difference*;
- *experience premium quality*;
- *designed for you*;
- *the perfect choice*;
- *unlock your potential*;
- *take your experience to the next level*.

Such expressions are grammatically well formed and persuasive, but repeated use across texts may reduce lexical dispersion. This distinction is important because AI-generated language can possess a relatively large vocabulary while simultaneously exhibiting high recurrence of particular promotional frames. Therefore, lexical diversity should be interpreted together with frequency distribution and lexical repetition.

English advertising texts. English human-written advertisements are likely to display considerable lexical compression. Because English has relatively limited morphological inflection compared with Uzbek, lexical variation can be observed relatively directly through changes in lexical stems and word forms.

Human-written English advertisements may use:

Fresh ideas. Bold flavors. A better way to live.

Such texts use short lexical units while creating semantic contrast through *fresh, bold,* and *better*.

AI-generated English advertising may produce a more explicitly descriptive lexical structure:

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Discover a fresh and innovative experience designed to bring exceptional quality, comfort, and convenience into your everyday life.

The second example contains more descriptive vocabulary, but it may also rely on common promotional collocations such as *fresh experience, innovative experience, exceptional quality,* and *everyday life*.

Consequently, AI-generated advertising may sometimes have higher lexical variety at the surface level while showing stronger dependence on conventional semantic frames.

Uzbek advertising texts. The analysis of Uzbek advertising requires special methodological attention because Uzbek is an agglutinative language. A single orthographic word may contain a lexical stem together with several grammatical or derivational morphemes.

For example:

mahsulotlarimizdan

can be segmented approximately into:

mahsulot + lar + imiz + dan

If lexical diversity is calculated solely on the basis of orthographic word forms, Uzbek texts may appear more diverse than texts in languages with different morphological structures.

Therefore, Uzbek lexical-diversity research should distinguish between:

- surface word forms;
- lemmas;
- lexical stems;
- morphological variants.

This distinction is particularly important when comparing human and AI-generated Uzbek advertising.

Human-written Uzbek advertisements may contain culturally specific expressions such as:

oilangiz uchun, *sifatli mahsulot*, *ishonchli tanlov*, *milliy ta'm*, *qulay narx*, *siz uchun*, and *mehr bilan tayyorlangan*.

AI-generated Uzbek advertisements may reproduce many of these expressions because they represent highly conventional promotional patterns. However, AI systems can also introduce English-derived or international advertising terminology such as *premium, digital, smart, online, modern,* and *innovative*. This suggests that lexical diversity in Uzbek advertising should be examined not only quantitatively but also from the perspective of cultural and morphological specificity.

Contrast between English and Uzbek. The contrastive dimension reveals several important differences. First, English advertising frequently uses compact lexical combinations and short slogan-like constructions. Uzbek advertising may employ more descriptive phrases, relational expressions, and culturally grounded vocabulary. Second, Uzbek morphology can produce multiple surface forms of the same lexical base. Consequently, lemmatization is particularly important when comparing lexical diversity across English and Uzbek. Third, borrowing and code-mixing may influence both languages. Contemporary advertising in Uzbekistan increasingly uses international marketing vocabulary such as *brand, premium, online, target, promo, cashback,* and *delivery*. Research on advertising vocabulary in Uzbekistan has documented the presence of English-derived technical and consumer-level terms as well as hybrid forms such as *targetlash* and *brending*. Fourth, cultural semantics can affect lexical selection. English advertising may emphasize individual choice and personal achievement, while Uzbek advertising may more frequently invoke family, trust, hospitality,

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

affordability, and collective values. Existing comparative research has identified culturally conditioned differences in persuasive strategies between English and Uzbek advertising.

Human versus AI lexical repetition. Lexical repetition is especially significant in advertising because repetition can enhance brand recognition.

However, excessive repetition may make AI-generated texts appear formulaic. For example, an AI-generated advertisement may repeatedly employ words such as:

quality – premium – innovative – unique – experience – perfect – discover

within a relatively short text.

Human writers may instead employ lexical substitution:

quality → excellence → craftsmanship → reliability

or:

beautiful → elegant → refined → sophisticated

Such synonymic variation can create greater lexical dispersion.

Nevertheless, repetition should not be interpreted negatively in all contexts. Strategic repetition of a brand name, product benefit, or slogan may be an intentional persuasive device. Therefore, the study distinguishes functional repetition from lexical redundancy.

Lexical fields. A comparison of lexical fields can further reveal differences between human and AI advertising.

Lexical field	**Human-written advertising**	**AI-generated advertising**
Quality	authentic, trusted, crafted, reliable	premium, exceptional, high-quality
Emotion	joy, comfort, confidence, delight	happiness, satisfaction, experience
Innovation	breakthrough, redesigned, next-generation	innovative, advanced, smart
Consumer	family, customer, community	you, your needs, your lifestyle
Action	try, taste, wear, explore	discover, experience, unlock
Value	affordable, worth it, everyday	excellent value, competitive, convenient

The table demonstrates that AI-generated advertising may favor semantically positive and broadly applicable vocabulary, whereas human-written texts can display more product-specific and context-dependent lexical choices.

AI and culturally specific lexical diversity. One of the most important questions concerns whether AI-generated advertising can reproduce culture-specific lexical diversity. In English, AI systems can relatively easily generate conventional advertising vocabulary because English advertising discourse is extensively represented in digital training data. In Uzbek, however, culturally specific advertising expressions may require greater contextual adaptation. For example, an Uzbek advertisement targeting family-oriented consumers may employ expressions related to *oila, farzand, ishonch, baraka,* or *mehmondo'stlik*. A literal translation of an English promotional message may fail to reproduce the pragmatic and cultural associations of such terms. Thus, lexical diversity should be considered together with lexical appropriateness and cultural specificity. Several limitations should be considered. First, advertising texts are generally short. This creates methodological difficulties for traditional lexical-diversity measures, particularly TTR. Length-sensitive measures should therefore be supplemented by MATTR and MTLD. Second, English and Uzbek have substantially different morphological structures. Direct comparison of surface word forms may therefore exaggerate differences in lexical diversity. Third, AI-generated texts depend heavily on prompts. Different prompts can produce substantially different lexical profiles. Consequently, future research should use

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

standardized prompts and multiple generations for each advertising task. Fourth, advertising genre itself influences lexical diversity. A slogan, product description, social-media advertisement, and long-form commercial text cannot necessarily be compared without controlling for genre and length. Finally, lexical diversity does not directly measure advertising effectiveness. A text with high lexical diversity is not necessarily more persuasive than a text with lower diversity.

CONCLUSION

The comparative analysis of human-written and AI-generated advertising texts in English and Uzbek demonstrates that lexical diversity is a multidimensional phenomenon shaped by authorship, language structure, cultural context, advertising genre, and communicative purpose. Human-written advertising tends to offer greater contextual specificity and may employ culturally motivated lexical choices, creative collocations, semantic variation, and controlled repetition. AI-generated advertising can demonstrate considerable lexical range but may also rely more strongly on conventional promotional vocabulary and recurrent advertising formulas. The English–Uzbek comparison reveals an additional methodological challenge associated with morphological typology. Because Uzbek is an agglutinative language, lexical diversity should ideally be measured at both surface-form and lemma/stem levels. This is particularly important when comparing AI-generated and human-written Uzbek advertisements. The study further demonstrates that TTR alone is insufficient for a rigorous comparison of advertising texts because of its sensitivity to text length. MATTR and MTLD provide more stable complementary measures and should be combined with frequency-based and qualitative analyses. From a practical perspective, AI-generated advertising should be evaluated not only according to grammatical correctness or vocabulary size but also according to lexical originality, cultural appropriateness, semantic precision, and control of repetition. For Uzbek advertising in particular, successful AI-generated content requires adaptation to local cultural and linguistic conventions rather than direct reproduction of generalized global advertising formulas. Future research should develop larger balanced English–Uzbek advertising corpora containing human-authored and AI-generated texts from the same product categories. Such corpora would make it possible to apply statistical significance testing, machine-learning classification, and multidimensional lexical analysis to determine which lexical features most reliably distinguish human and AI-generated advertising discourse.

References

1. McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392.
2. Melnyk, O., & Kyselova, I. (2025). Lexical neologisms in English: Formation, trends, and cultural impact. *Studia Philologica*, 24(1).
3. Plag, I. (2003). *Word-Formation in English*. Cambridge: Cambridge University Press.
4. Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
5. Ruzmuratova, A. A. (2025). The formation and sources of neologisms in modern Uzbek and English lexicons. *Modern American Journal of Linguistics, Education and Pedagogy*, 1(3), 710–714.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

6. Azimova, D. A., & Yusupova, S. B. (2025). Contrastive analysis in advertising discourse in Uzbek and English languages. *American Journal of Philological Sciences*, 5(6). DOI: 10.37547/ajps/Volume05Issue06-04.
7. Khusanova, U. (2025). Linguistic analysis of advertising texts in English and Uzbek. *International Conference on Science, Education & Law*, 1(3), 9–12. DOI: 10.5281/zenodo.17862701.
8. Kim, T. (2026). Introducing and evaluating a measure of lexical diversity across word classes. *TESOL Quarterly*. DOI: 10.1002/tesq.70154.
9. Kendro, S. (2026). Do Large Language Models Produce Texts With "Human-Like" Lexical Diversity? Evidence From Four ChatGPT Models. *International Journal of Applied Linguistics*. DOI: 10.1111/ijal.70115.
10. Sobirova, M. (2026). The role of verbal and non-verbal components of advertising in English and Uzbek languages: A comparative multimodal study.