



COLLOCATION PATTERNS IN PAKISTANI BILLBOARD ADVERTISEMENTS: A CORPUS-BASED ANALYSIS

Maham Shuaib

mahamshuaib5@gmail.com

MPhil Scholar, Department of English, University of Okara, Pakistan.

Muhammad Kamran Abbas Ismail

Kamran.ch@uo.edu.pk

Lecturer, Department of English, University of Okara, Pakistan.

Mishal Abbas

mishalabbas947@gmail.com

MPhil Scholar, Department of English, University of Okara, Pakistan.

Mubara Ghani

mubaraghani9@gmail.com

MPhil Scholar, Department of English, University of Okara, Pakistan.

Abstract

The purpose of this study is to provide the first corpus-based study of collocation patterns in billboard advertisements in Pakistan. A corpus of 70+ unique billboard texts (793 tokens) was created for the four major cities of Lahore, Karachi, Islamabad and Multan across the years 2021 to 2025. The high-frequency words were analysed in LancsBox v5.5.1 using a 5L/5R window for Mutual Information (MI) scores. The lexical associations were very strong, as demonstrated by all the collocates, with three main discourse strategies emerging: promotional (free, limited, offers), claims of quality and improvement (new, best, upgrade), and action-oriented imperatives (get, available). The results offer a new MI-based collocation frequency list for advertising English in the Pakistani context, with implications for the development of advertising copy, English instruction in Pakistan and the localisation of worldwide marketing strategies in South Asian contexts.

Keywords: *Advertising Discourse, Billboard Advertising, Collocation Analysis, Corpus Linguistics, Outdoor Advertising, Pakistani English.*

Corresponding Author: Muhammad Kamran Abbas Ismail (Lecturer, Department of English, University of Okara, Pakistan)

Email: Kamran.ch@uo.edu.pk

1. Introduction

1.1. Background

Billboard advertising is an important part of the marketing mix in urban areas of Pakistan, where literacy and the prevalence of cars make billboards accessible. With Pakistan's economy becoming more active in global markets, outdoor advertising has shifted from conveying product information to a multifaceted linguistic structure that builds brand image and triggers consumer action. Despite their prominence and cultural significance in Pakistani cities, billboards are rarely empirically researched systematically in Pakistan.

Corpus-based methods provide a very useful quantitative tool for analysing naturally occurring data in specialised discourse communities. Collocation analysis (the statistical discovery of word co-occurrence patterns) has been particularly useful for uncovering the conventional and/or persuasive aspects of language use across registers. This is encapsulated in Firth's (1957) maxim "you shall know a word by the company it keeps". Such analysis can provide not only an understanding of the words advertisers use but also the lexical pairings they regularly use to build meaning, engage consumers, and set themselves apart in the competitive outdoor advertising environment.

The gaps in the literature this study seeks to fill are interrelated and formulated as follows: To the best of the author's knowledge, there is no systematic corpus study of collocation patterns in Pakistani billboard advertisements, though the register has high visibility and influence in the urban public sphere. While there has been some research on Pakistani English in newspaper discourse, social media and educational settings, there has been virtually no study of this field in outdoor advertising. Second, there has been a focus on Western markets, particularly the U.S. and U.K., yielding geocentric knowledge that is not immediately transferable to multilingual South Asian environments with distinct cultural values, consumer behaviours and variants of English. Third, a comparison of collocation approaches across product sectors (telecommunications, fashion, food and education) in Pakistani OOHM has not been done, so that the sector-specific issues of collocation of advertising language have not been explored. Fourth, there is no corpus-based study for the use of urgency and promotion terms (free, limited, upgrade) in Pakistani billboards, which also play a vital part in OOHA discourse. Lastly, there is a lack of an MI-based collocation frequency list (CFL) of Pakistani advertising English for the benefit of copywriters, language teachers and World Englishes (WE) researchers, to be used for practical and scholarly purposes.

1.2. Research Questions

The following are the research questions in this study:

1. What are the most common co-occurrence relations between the high-frequency promotional words in Pakistani billboard advertisements?
2. How do these patterns illustrate persuasive discourse patterns and conventions of Pakistani advertising?

3. What are the unique collocational patterns in the use of English in the commercial context in Pakistan?
4. What is the difference between the collocations found in the identified texts and those in the texts from Western advertising?

1.3. Significance of Study

This study is a contribution to corpus linguistics because it uses a well-known methodology and applies it to a new register and geographical region. It offers empirical evidence of linguistic strategies in use in a key South Asian market and challenges Western-centric models applied to the study of advertising. It is a work of World Englishes scholarship that establishes specific lexical traits of Pakistani English, thereby fostering the understanding of how global languages are locally adapted. It also produces authentic and corpus-derived texts that reflect the English used by Pakistanis in their urban lives. The article is organised into the following sections: Section 2 reviews the related literature; Section 3 describes the methodology; Section 4 presents the collocation results; Section 5 discusses the findings in relation to the research gaps and comparative perspectives; and Section 6 concludes with limitations and future directions.

2. Literature Review

2.1. Corpus Linguistics and Collocation Theory

The methodological paradigm of the mid-20th century that coalesced into corpus linguistics is grounded in patterns of naturally occurring language and a focus on meaning. In Firth's (1957) work, the contextualist approach laid the foundation for understanding that meaning is not in the meaning of lexical items but in the co-occurrence relations between them. This insight was taken further by Sinclair (1991) in a detailed analysis of a corpus, where he showed that language users predominantly use prefabricated phrases and conventional word combinations, not just generative grammar, and that these are governed by what he called the idiom principle. Stubbs (2009) described this as the universal repetition of familiar units, and it has important consequences for language acquisition, processing and register variation.

Theoretical and methodological aspects of collocation research have been further elaborated in the following scholarship. Bartsch and Evert (2014) expanded the Firthian view of collocation, focusing on frequency, mutual expectancy and semantic non-compositionality as the hallmarks of true collocation. Lehecka (2015) identified two types of collocational phenomenon: lexical and grammatical; whereas the former is included in the latter, both can be part of the wider spectrum of phraseological phenomena in language. Multi-word units, made up of collocations, colligations and semantic prosody, have been shown by Phoocharoensil (2021) to have meanings greater than the sum of their parts and to carry attitudinal and evaluative elements as well as their compositional semantics. Through large-scale corpus studies, Biber et al. (1998) have demonstrated that statistical association measures can effectively distinguish genuine collocations from random co-occurrence, thus making

corpus-driven studies particularly useful in identifying emergent patterns that exist in a corpus but would likely go unnoticed in the context of a single person.

Collocation studies employ a variety of statistical tools. Several statistical measures are used in collocation studies.

Measures of association strength are needed to identify collocations. Church and Hanks (1990) proposed a principled definition of the degree to which two words co-occur beyond chance: $MI(x,y) = \log_2[P(x,y) / (P(x) \times P(y))]$, where $P(x,y)$ is the probability of the words co-occurring and $P(x)$, $P(y)$ are the probabilities of each word individually. The higher the MI score, the greater the lexical association. Evert (2005) outlined a detailed mathematical analysis of word co-occurrence statistics and noted that MI is especially sensitive to exclusive collocations, which consist of co-occurring word pairs that occur consistently but rarely in other combinations. Kolesnikova (2016) reviewed the most common methods, namely, frequency-based, probability-based, information-theoretic and hypothesis-testing approaches, and she concluded that there is no universal method that is clearly better, the choice depending on the research objectives and the nature of the corpus. In specialised corpora, Spicer (2022) noted the need to consider the context of other language elements when applying MI, including minimum frequency thresholds and window span selection. This is an important point made by Deng et al. (2022), who demonstrated that even in small, specialised corpora, it is possible to find domain-specific lexical patterns that are absent from general reference corpora, which is directly relevant to the present study. In a similar vein, Wermter (2009) showed that small, purpose-designed corpora can reveal domain-specific lexical relationships that do not stand out in general reference collections, thereby affirming the approach adopted in this work.

Advertising discourse has been a focal point of research interest as both a social phenomenon and a language type, given its universal presence in society and its specific linguistic characteristics. While most work is qualitative and based on small samples, Teodorescu (2015) found that a number of strategies, such as repetition, parallelism, wordplay and imperatives, are characteristic of advertising messages. The corpus approaches have started to reveal patterns not otherwise provided by discourse approaches: Ha (2023) discussed the semantic distinction in advertising lexis; Aihua (2013) explored high-frequency lexical and grammatical features of advertising English; Gugić (2021) has documented adjective-noun collocational patterns in tourist advertising brochures; and Lakkam and Manowong (2020) have manifested product-specific collocational profiles in skincare advertising. Hafeez et al. (2025) used corpus analysis in the Pakistani context of digital ads. This is accounted for by cross-cultural comparisons, such as those used by Matte and Lenz (2019), who compared academic writing collocational preferences in Brazilian and British varieties, showing that collocational preferences have definite cultural specificity and, therefore, the same would apply to advertising language. The multimodal corpus analysis tools created by Todd et al. (2023) indicate a broadening of the methodological scope of advertising. Additionally, Sun et al. (2022) delved into the application of natural language processing methods for customer opinion

analysis, highlighting the evolving relationship between corpus linguistics and computational consumer research tools.

2.2. Pakistani English and Advertising Discourse

Pakistani English has emerged as a unique variant of the World Englishes system, with phonological, lexical, grammatical and pragmatic features that reflect Pakistan's multilingual environment. Pakistani public signage is a complex area of code-mixing and language choice, documented by Manan et al. (2021), and coexists with Urdu and regional languages. Previous studies of Pakistani English have increasingly identified a range of features, including lexico-semantic (Jilani et al., 2018), lexical (Amir et al., 2025), and syntactic (Imran, 2023). All these studies agree that Pakistani English is not a deviant form of native-speaker English but a distinct variety with its own systematic and well-documented characteristics.

Most research in Pakistan focusing on Advertising has been devoted to the study of code-switching and code-mixing. Riaz (2019) studied strategic switching between English and Urdu in TV commercials; Shah et al. (2025) analysed code-mixing in social media ads; Habib et al. (2025) studied code-switching in Pakistani billboard ads; and Munir et al. (2025) examined the persuasive aspects of code-switching in taglines. The investigation of language use in monolingual English billboard advertising has not been well explored, as it is the subject of the present study. Meanwhile, Shah et al. (2019) used critical discourse analysis to analyse Pakistani billboards without a corpus-based methodological approach and without consideration of the collocational approach. Awan et al. (2025) used a corpus-assisted methodology with a methodological precedent in the field of educational print advertisements, and Sardar et al. (2020) and Anwar et al. (2025) identified lexical variation in advertising in Pakistan. The present study continues this approach and extends corpus-based approaches to the register of English-dominant outdoor billboard advertising, which has not been explored before, especially with regard to the use of MI-based collocation analysis. It contributes to Pakistani English literature and the wider field of South Asian advertising discourse, thereby bridging an empirical gap.

3. Methodology

3.1. Corpus Compilation

The Pakistani Billboard Advertisement Corpus was created by systematically photographing outdoor advertising displays in four major cities of Pakistan, namely Islamabad, Lahore, Karachi and Multan. These cities have been chosen to represent Pakistan geographically, economically and culturally. The commercial and cultural capital of Punjab, Lahore, offers significant advertising potential, with high commercial concentration along major arterial roads. The economic hub of Karachi also hosts the headquarters of national and multinational companies, with high advertising investments. Islamabad's planned urban design and federal status create a unique context for advertising, while Multan illustrates the dynamics of the southern market in the province of Punjab, reflecting the experience of a big commercial hub outside the main metropolitan hub.

The data were collected from 2021 to 2025 to capture the current state of billboard advertising in Pakistan, during a period of massive economic and technological transformation. Photographs of billboards were taken in high-volume commercial areas, major roadways, and shopping areas. Only English-dominant billboard texts (either monolingual English or English-dominant bilingual) were considered as part of the corpus. Billboards that were completely in Urdu or regional language and/or image-heavy with little text were ineligible, while mixed language billboards where the English language was the primary or predominant language were eligible. This criterion has been designed to capture the register of commercial English used in outdoor advertising in Pakistan, which frequently features English-dominant text that may be accompanied by Urdu taglines or product names, but does not disrupt the predominantly English collocational environment of the outdoor advertising medium.



Figure 1. Billboard Advertisement

The resulting corpus consists of 793 tokens from 70 unique billboard advertisements, similar in size to those used in previous specialised advertising discourse research (e.g., Lakkam & Manowong, 2020). Photographs were read several times for transcription, with original spelling, capitalisation, and punctuation maintained, and the transcription was done by hand. The collection covers a variety of product areas, including telecommunications (Telenor, Zong), automotive services (CarFirst), fashion and retail (Zellbury), food and beverages (McDonald's, various local brands), real estate, education and consumer goods (Diamond Supreme Foam, earbuds (Ronin), Slanty), allowing for cross-sector collocational comparisons. The size, location (roadside versus building-mounted), and type of the advertiser (multinational corporation versus local business) of the billboards also show variation, making the corpus representative of the diversity of Pakistani outdoor advertising practice.

Table 1: Corpus Distribution by Product Category

Product Category	Representative Brands / Examples	Approx. % of corpus
Telecommunications	Telenor, Zong	~22%
Fashion & Retail	Zellbury, various retailers	~18%
Food & Beverages	McDonald's, local food brands	~17%
Consumer Goods	Diamond Supreme Foam, Ronin, Slanty	~15%
Automotive Services	CarFirst	~10%
Real Estate	Various developers	~9%
Education	Various institutes	~9%

Table 1 shows the distribution of billboard texts by product category in the corpus, which reflects the sectoral distribution of the data.

The use of data processing and analysis software.

Corpus analysis was conducted using LancsBox X 5.5.1 (Brezina et al., 2015), which offers powerful extraction and visualisation of collocations through its GraphColl functionality. LancsBox is a freely available, widely validated corpus analysis platform widely used in corpus-based discourse studies. Minimal pre-processing to retain the “true” nature of advertising language: typographical errors corrected, but non-standard creative spellings left as an intentional choice; text lowercased for frequency and collocation analysis; original capitalised in source files for qualitative reference. In the source documents, proper nouns and brand names were left as they were and treated as lowercase during quantitative analyses, so the frequency of each was counted—this was done to be consistent across variant capitalisations.

5 words left (L5) and 5 words right (R5) was used as a window frame — a commonly used window frame in corpus collocation studies that is able to cover a sufficient number of lexical associations, yet excludes contextually distant words, and that may not have a meaningful relationship with the node word. The main association measure was Mutual Information (MI), with a minimum collocate frequency of 1. Using a higher minimum

frequency would result in the loss of potentially meaningful collocates that appeared only in the advertising register, not in the general register.

3.2. Node Word Selection

The words were chosen from the top 10 words in the corpus and were also relevant for the discourse function in advertisements: free, new, Pakistan, available, best, get, upgrade, limited, offers, your. These terms include together: promotional terms (free, limited, offers); quality and improvement terms (new, best, upgrade); action and access terms (get, available); geographical identity terms (Pakistan); and direct address terms (your). It includes a wide variety of grammatical categories, such as adjectives (free, new, best, limited), verbs (get, upgrade), nouns (offers), determiners (your) and a proper noun (Pakistan), thus providing a wide spectrum of lexical patterning of the billboard. The presence of an appropriate geographical identity (Pakistan) as well as functional advertising vocabulary indicates the study's interest in documenting the advertising discourse patterns found in general and the locally specific features of Pakistani commercial English.

3.3. Statistical Measures and Interpretation

MI was operationalised as $MI(x,y) = \log_2[P(x,y) / (P(x) \times P(y))]$. A score of 3 or higher is a true collocation, and a score of 3–5 represents some association, while a score of 5 or higher signifies strong association. The MI scores for all collocates in the present study ranged from 5.5 to 9.6, all very high. This uniformity is typical of the highly conventionalised nature of billboard advertising discourse, with the limited number of acceptable word combinations yielding high MI scores in a small, specialised corpus. The upper limit of 9.6 (upgrade–not) is one of the highest MI scores reported in advertising collocation studies and represents an extraordinarily close collocational bond created by the fixed formula: upgrade your sleep, not your expenses. It is especially suited to MI use because the aim of the exercise is to typologize Pakistanis' billboard advertising language rather than general Pakistani English.

Quantitative MI analysis and qualitative discourse analysis of semantic and pragmatic functions were conducted to generate interpretation, following the mixed-methods approach recommended by Spicer (2022) and Deng et al. (2022). The semantic and pragmatic functions of collocates in the formation of persuasive ads, supported by authentic billboard samples from the corpus, were analysed; in addition, the statistical relationships were reported. This integration allowed us to reach a new level of understanding of how collocational patterns contribute to advertising discourse and to go beyond statistical description to consider the functional and communicative motivations of the observed collocations.

4. Results

This unit presents an overview of collocation patterns.

The analysis of the words on the nodes of the LancsBox revealed very specific lexical links for each of the ten node words, with MI values ranging from 5.5 to 9.6. A comprehensive summary is given in Table 1. With a mean MI score of about 7.3, the overall score of the ten

node words shows that advertising in Pakistani billboards has strong collocational stability, which is higher than the threshold for genuine collocation ($MI > 3$) and is in the upper range of the scores reported in advertising discourse studies (cf. Lakkam & Manowong, 2020). The results are found to be interesting when the distribution of MI scores for various semantic categories of node words is considered, such as urgency and scarcity terms, which get highest scores in their conventionalised collocational pattern in Pakistani billboard ads (limited: 8.0; offers: 8.0–8.6), as well as lifestyle-improvement words, which also get the highest scores (upgrade: 7.3–9.6).

Table 2: Collocation Patterns in Pakistani Billboard Advertisements (#LancsBox, MI Scores)

Node Word	Top Collocates	MI Score Range
Free	Spray, pillows, sound, collectibles, body, exclusive, creativity, celebrate	7.3
New	Dream, magic, taste, unstoppable, zong	6.5–7.0
Pakistan	Billboard, drone, McDonald's, safety, days, biryani, chips	5.6–7.2
Available	Online, stores, sound, collectibles, creativity, free	6.7–8.0
Best	Offers, look, shop, arrivals, friday, just	8.0
Get	Digital, let's, 4g, front, customer, updated, zong	5.7–6.8
Upgrade	Not, convenient, comfortable, expenses, sleep, your	7.3–9.6
Limited	Slanty, earbuds, edition, sound, stickers, pillows, exclusive	8.0
Offers	Best, friday, just, upto, look	8.0–8.6
Your	Upgrade, expenses, not, sleep, limelight, consultant	6.8–7.8

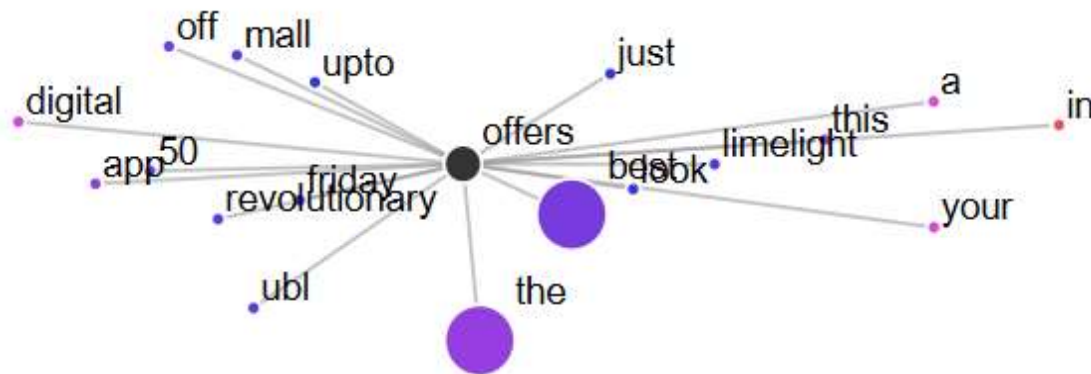


Figure 2 Graphcoll visualization for node word “offers”

The GraphColl visualization shows that upgrade frequently co-occurs with words related to improvement and lifestyle enhancement. This highlights the use of positive change as a persuasive advertising strategy.

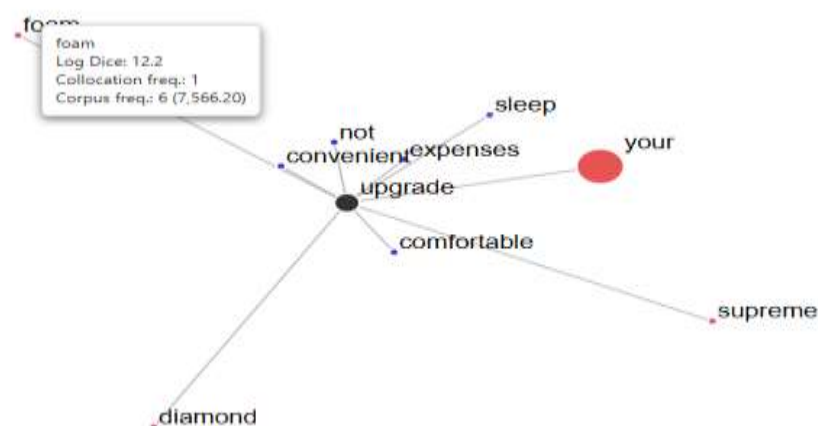


Figure 3 Graphcoll visualization for node word “upgrade”

The GraphColl visualization indicates strong associations between offers and promotional vocabulary. This reflects the importance of discounts in Pakistani billboard advertising.

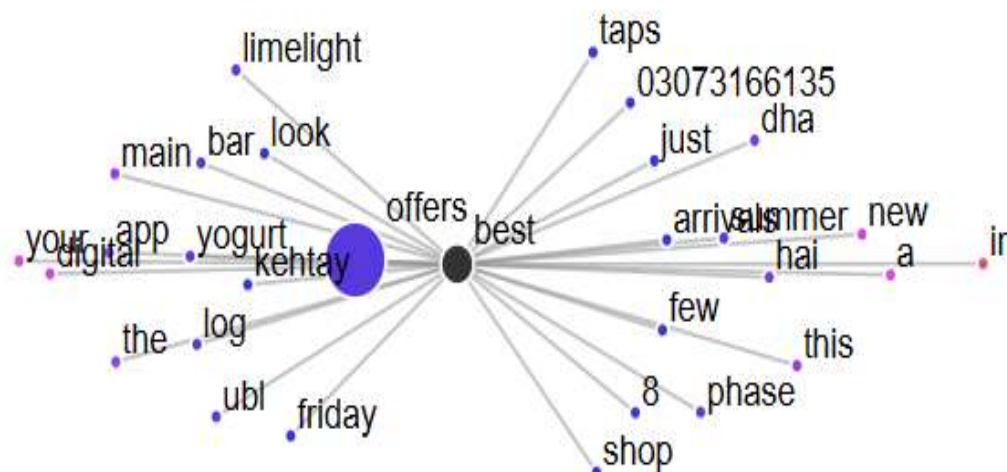


Figure 4 Graphcoll visualization for node word “best”

The GraphColl Visualization demonstrates that best commonly collocates with words expressing quality and superiority. This suggests that advertisers use evaluative language to strengthen product appeal.

4.1.1. Free

Free associates with tangible giveaway items, such as spray (7.3), pillows (7.3), sound collectibles (7.3) and body (7.3); as well as experiential enticements like exclusive (7.3) and celebrate (7.3). The pattern is to couple free with transactional incentives (free body spray, free pillows, free sound collectibles) and lifestyle-oriented appeals (feel free to celebrate your creativity). The MI of ‘free’ across various collocates suggests that ‘free’ has a stable, multifunctional role in Pakistani billboard discourse: it constructs value by promising material benefits from its supplementary part, and it also signals emotional freedom for authentic self-expression. The presence of free with products like pillows and sound collectibles, items that are tangible and physical, and have a lower actual cost but higher perceived value in the eyes of the consumer, is consistent with the “free gift with purchase” strategy that is common in Pakistan’s retail sector, where consumers are price-sensitive and anything that is tangible and physical, such as a pillow, is a good way to incentivise them to buy.

4.1.2. New

New collocates with dream (7.0), magic (7.0), taste (7.0), unstoppable (7.0), and the telecom brand zong (6.5). These associations frame novelty as transformative and aspirational: new products offer new experiences, as Zong does in the telecommunications industry through its brand narrative of network expansion and upgraded digital services. Collocate taste (7.0)

suggests food and drinks, with new flavour launches becoming a major topic on billboards in urban areas in Pakistan. While the collocate unstoppable (7.0) seems to have a similar meaning to an adverbial, it is more aspirational, typical of the register of the telecom and lifestyle ad campaign, in which the product's novelty is seen as a means of personal empowerment and permanent progress. These co-occurrence collocates suggest that 'new' in billboard advertisements in Pakistan serves two roles: product description and aspiration trigger.

4.1.3. Pakistan

Pakistan appears with billboard (5.9), drone (7.2), McDonald's (6.2), safety (5.6), biryani (5.6), and chips (6.2). The high MI for drone (7.2) is due to the coverage Coca-Cola's drone billboard campaign generated for the technology-driven billboard advertising concept, a concept that has attracted significant media and advertising industry attention in Pakistan and is therefore frequently referenced in billboard-related texts. The co-occurrence of Pakistan with food terms biryani (5.6) and chips (6.2) indicates that food and quick-service restaurant advertising is a strong feature of Pakistani urban outdoor media, with food that resonates with the local culture and are culturally significant (e.g. Pakistan's biryani and chips association with McDonald's), localising the messaging of brands. The collocate safety (5.6) is related to public service and corporate social responsibility advertising, suggesting that billboard advertising in Pakistan involves both civic (non-commercial) communication and commercial promotion. It is the function of the billboard corpus as a place of commerce and as a site of national identity-building, as revealed by the node word 'Pakistan'.

4.1.4. Available

Known by its connection to available (7.3), stores (7.0), sound (8.0), collectibles (8.0), creativity (8.0) and free (6.7). The phrase "available in stores & online" recurs in their ads and has become commonplace in Pakistani advertising, as e-commerce expands in tandem with traditional brick-and-mortar retail channels. The particularly high MI for sound (8.0) and collectibles (8.0) indicates a specific product-category cluster in the corpus with a significant focus on audio products and limited-edition merchandise (in particular Ronin earbuds, as well as associated branded collectibles) where availability messages are especially important because the scarcity is both a logistical reality and a marketing opportunity. When available co-occurs with free (6.7), a phrase like 'available with a free gift' suggests a strong relationship between available and promotional value in Pakistani billboard discourse.

4.1.5. Best

Top frequencies are offers (8.0), look (8.0), shop (8.0), arrivals (8.0), and friday (8.0), with quality claims superlative and used in conjunction with promotions, appearance, and new product launches. The MI of 8.0 is a very interesting finding, reflecting the very stable collocational environment of "best" in billboard advertising in Pakistan. The best offers combination is the most commercially direct since it positions the advertiser's offers as objectively better than the others. Best-look and best-arrivals are examples of the fashion and retail industries' use of 'best' in the superlative sense as an aesthetic positioning and a new-

stock promotion, respectively. The collocate Friday (8.0) added to offers confirms the framing of the promotion event discussed for offers: claims of quality are coupled with the time-limited promotion event Friday. Across all these collocates, the uniform MI of 8.0 indicates that quality positioning (in the superlative form best) is one of the most conventionalised and stable patterns in billboard advertising in Pakistan.

4.1.6. Get

Get is an action imperative, like digital, 4 G, updated, let's, and the brand Zong. These correspond to the predominance of telecommunications advertising in the corpus and the sector's need to call for action in consumer behaviour. The collocate digital (6.2) places "get" within the wider narrative about Pakistan's digital transition, in which telecom advertisers position their services as vital digital infrastructure for participation in the modern digital economy. The phrasing of 'get updated' (6.8) suggests that the lack of the advertised service is a lack of information and/or technology, not a product issue. The recurrent phrase "let's get..." (6.8) adds a participatory, community component to the call to action, making consumer behaviour a social process rather than just a transaction.

4.1.7. Upgrade

The combination of the study's highest MI rating (9.6) and not comes from the billboard campaign by Diamond Supreme Foam's slogan "upgrade your sleep, not your expenses. The phrase is a key aspect of Pakistan's advertising industry's marketing strategy, which here promises quality enhancement and cost reassurance, tailored to suit price-conscious Pakistani consumers. The collocates convenient (9.6) and comfortable (9.6) occur closely together in the sphere of physical ease and lifestyle improvement, suggesting the change is easy and inexpensive. The term is well grounded in the bedding/home goods area, thanks to its collocate sleep (7.8). The collocation your-upgrade repeats this node word, and provides a personalised consumer-brand relationship through the prospect of individualised improvement, as the possessive address strategy is analysed under the node word your. The only node word in the corpus that is semantically dense and strategically sophisticated is upgrade.

4.1.8. Limited

Limited collocates with slanty (8.0), earbuds (8.0), edition (8.0), sound (8.0), stickers (8.0), and pillows (8.0). The uniform MI of 8.0 is typically matched with product-specific scarce objects (PSSO) in Pakistani billboards, resulting in product-specific scarce appeals for collectible, exclusive physical objects. The collocates slanty (a Pakistani snack brand) and earbuds (Ronin audio products) suggest the use of promotional campaigns with limited-edition co-branded and special-run products that create a sense of urgency to purchase, leveraging the concepts of brand loyalty and scarcity psychology. The collocate stickers are particularly interesting, as they show that even a low-cost supplementary item (collectible stickers packaged with products) is used in a scarcity discourse in Pakistani billboard advertising, thus revealing the sensitivity of Pakistani consumers to collectibility appeal. Collocate: The exclusive (8.0) adds to the high-quality positioning of limited-edition products.

4.1.9. Offers

Renders special, time-limited value offers to associates, such as best (8.0), friday (8.6), just (8.6), and upto (8.6). The MI score for Friday is among the highest in the corpus for this node word, indicating that “Black Friday” has now become a cultural and commercial phenomenon in the Pakistani OOH industry. This phenomenon has been stripped of its American cultural context and made a universal marketing signifier in Pakistani markets. The formula of the advertising discourse used in the sentence “offers upto 50%” is important as a uniquely Pakistani English construction; the preposition upto (not upto 50% as in British and American English, but upto 50%) is a non-standard preposition that is used as a conventional quantifier in the advertising discourse. The collocation just (8.6) in the context of ‘just for you’ personalises the offer, linking the promotional discourse of offers to the modality of the direct strategy through the node word your.

4.1.10. Your

Your collocates with upgrade (7.8), expenses (7.8), sleep (6.8), limelight (6.8), and consultant (6.8). The possessive pronoun is an active call to personal space, to lifestyle improvement, money matters and career aspirations, bringing consumers into their personal space. Your-upgrade (7.8) links this word to the lifestyle-enhancement discourse associated with the upgrade node, while your-expenses (7.8) directly speaks to financial concerns, making the consumer price-sensitive and reassuring. The collocate limelight (6.8) is an example of a fashion and beauty way of creating an aspirational consumer self-image: “to be in the limelight”. It comes from a direct flattery and social aspiration strategy. One of the most distinctive aspects of billboard advertising in Pakistan is collocate advertising, which is especially associated with education and career consultancy advertising and establishes a relationship between the billboard consultant and the public by directly addressing them, with the billboard serving as the vehicle for conveying personalised knowledge. This pattern does not have a clear counterpart in western advertising corpora and is among the most distinctly Pakistan-specific collocational patterns identified in the study.

5. Discussion

5.1. Addressing Research Gaps

5.1.1. First Corpus-Based Study of Pakistani Billboard Collocations.

This study is the first systematic, corpus-based study on collocation patterns in Pakistani billboard advertisements. The use of corpus-based approaches to advertising discourse in the West has been attempted before (e.g., Aihua, 2013; Gugić, 2021; Lakkam & Manowong, 2020), but there has been no previous work on corpus-based approaches to Pakistani outdoor advertising discourse in English. The high MI scores (5.5–9.6) obtained for all node words support the general idea of a high degree of lexical conventionalisation in the discourse field, in line with the principle of idiosyncrasy derived from Sinclair (1991) which stipulates that prefabricated, predictable word combinations determine language use. The finding is also corroborated by Biber et al. (1998), who showed that statistical association

measures are reliable for differentiating genuine collocations from random co-occurrence in their corpus-based framework. In the context of Pakistani educational advertisements, Awan et al. (2025) conducted a corpus-assisted analysis, which identified systematic lexical patterns that further corroborated the validity of the corpus-based approach to this under-researched South Asian register.

5.1.2. The Need for Advertising Research to Go Beyond Western-Centric Approach to Advertising

The results oppose Western-centricity by presenting collocation patterns in a major market with unique linguistic, cultural and economic features. There are universal promotion strategies (free, new, best) and context-specific collocates of these node words (constructs value perceptions across cultures). The collocates in this data set (change, consultant) reflect the national development discourse and the unique professional services markets in Pakistan. The local market structure is indicated by the Brand collocates (Carfirst, Vangli, Telenor). The glocalisation of the global phenomenon of outdoor retail space is illustrated in the collocation “offers friday”, as discussed by Manan et al. (2021) in relation to English in the Pakistani landscape. Similarly, Trung et al. (2025) noted the same urgency strategies in out-of-home advertising media across Asia and placed Pakistani patterns within regional trends. The present study shows that there is greater emphasis on transactional and promotional language (get, offers, available) in the tourist advertisements because of the short communication length of billboards and the price-sensitive consumer environment in Pakistan, compared to Gugić (2021), which revealed a high proportion of the evaluative adjective-noun combinations in the tourist advertising, which reflected aesthetic and experiential aspects.

5.1.3. Do comparisons of collocation strategies cross categories

This is the first study to compare collocation strategies across two different categories of out-of-home media; the results reveal common and sector-specific patterns. In the case of telecommunications advertising (get telenor, get weekly), the emphasis is on the availability of service and recurring benefits characteristic of a subscription-based business model, in which the key consumer behaviour is signing up for a service rather than purchasing something once. Automotive services (e.g., Carfirst) are well-suited to action-based transactional language, which allows consumers to take action in a single step. As fashion consumption is experience- and identity-oriented, novelty and quality positioning (new arrivals, best look) are highlighted in fashion and retail advertising. Marketing home goods and furniture creates very product-specific clusters (pillows, sleep, comfort, upgrade), which refer to experiential properties and qualities of physical home goods.

Although there are differences across categories, some discourse strategies cut across sectors. The use of urgency and promotional language (free, limited-time offers) was observed across all product categories in the corpus, indicating that such appeals are common and effective in the Pakistani market. Quality and improvement claims (best, new, upgrade) span across categories, suggesting a somewhat similar underlying motivation among consumers to

obtain the best value and keep up with the times. The cross-category analysis has practical implications for advertising professionals in Pakistan as knowledge of which collocations constitute an established formula for a specific sector and which are universal can inform the copywriting process, and help the advertiser either use the established formulas of the sector to ensure the ad will be recognised, or introduce patterns from other categories to the ad in order to provide novelty in an already competitive outdoor advertising market.

5.1.4. Analysis of Urgency and Promotional Language.

The ten node words include three (free, limited, offers) that express direct promotional appeals and two that promote transactional urgency (get, available), which make up the bulk of the high frequency node words in the corpus. Combined with physical giveaway items (spray, pillows, sound collectibles), Free, plus free items given away with purchase (free gifts with purchase) reduces perceived purchase risk and increases perceived value, especially in Pakistan's price-sensitive retail market, where additional free items can be key factors in consumer buying decisions. Limited, collocating with concrete, specific objects – slanty stickers, Ronin earbuds, pillows – establishes product-specific scarcity based on the psychology of limited availability. In corpus examples, such as 'Limited offer available', the temporal dimension of scarcity is made explicit through both the accessibility (current) and the implied time pressure. Collocates offers with temporal (Friday), quantitative (upto), and evaluative (best) terms and describes promotional events as exceptional, bounded opportunities.

"offers Friday" is a common occurrence in the corpus and an important process of cultural localisation, as globally popular retail events like Black Friday are imported to Pakistani markets and used to create a recognised retail event in outdoor advertising, with the original cultural context removed and promoted as an event. The urgency and promotional language in the billboards are the result of the competitive retail environment, the price consciousness of Pakistani consumers, and the limited attention span of billboard advertising, which demands short, easy-to-understand promotional messages. The promotion language is found in most of the billboards across Pakistan, and the collocations (free pillows, limited stickers, upto 50%) are unique to the Pakistani market, in line with Teodorescu (2015), who found that creation of urgency was a universal advertising strategy, and to the Pakistani English L-G feature (upto as a single-word quantifier), which could be further investigated.

The first collocation list based on dictionaries and MIs for Advertising English for Pakistani students is 5.1.5.

This study provides the first MI-based collocation frequency list for Pakistani advertising English, which goes beyond the newspaper register (Jilani et al., 2018; Amir et al., 2025) and documents the commercial outdoor advertising register. For corpus linguists, the list reports on English-language features found in a register that has not been described or investigated before, adding to the growing empirical evidence for characterising the features of Pakistani English as systematic and documentable. Previous corpus-based studies of

Pakistani English had limited their scope to the language of newspapers and neglected commercial registers to a large extent. The present study generalises this documentation to the outdoor advertising domain, where it is shown that the collocational features of Pakistani English in the commercial versus news domains differ.

The collocation list offers valuable insights for copywriters and marketers in Pakistan, helping them craft effective advertising messages based on data rather than on the advertising models used in Western countries. The high MI scores indicate that the collocations identified are not only frequent but also conventional in Pakistani billboard discourse, suggesting that Pakistani consumers are likely to process them fluently and relate them to familiar advertising genres. From an English language teaching perspective, the list provides authentic resources for English as used in the city (as Jablonkai (2020) argued in this article for productive vocabulary knowledge), which is consistent with the article's overall theme. Collocations like Pakistan consultant, get Carfirst, and new Vangli are quantitative evidence of the systematic localisation in Pakistani Commercial English for world Englishes scholars. The MI-based approach does not focus solely on the most common word combinations, but also on those that are statistically significant, which is important for capturing the truly conventional word combinations characteristic of this specialised discourse register.

5.2. Comparative Perspectives

Western advertising research terminology is compared with Pakistani billboard collocations, and similarities and differences are identified. In line with Matte and Lenz (2019), collocational variation results from the interaction between local communicative practices and universal persuasive principles. The abundance of promotional language (free, offers, limited) validates Teodorescu's (2015) finding that value-oriented appeals are a constant in European advertising. Aspirational and lifestyle terms such as 'new unstoppable lifestyle' and 'upgrade lifestyle' are also found on Pakistani billboards, but more often than not in combination with more transactional terms than is the case in the West; this can be attributed to a price-sensitive consumer market in Pakistan and the immediate communication requirements of outdoor advertising.

References to geographical and national identity are especially prevalent in Pakistani billboards than in Western advertising literature. The code-switching of Pakistan and safety through the use of drones and billboards, along with days, demonstrates a unique kind of technological modernity and civic responsibility that is seldom reported in the Western corpus-based advertising literature. This could reflect a nation-building narrative in a developing economy, where national identity is a significant marketing advantage and the need to make global brands meaningful to Pakistani identity. As previously reported by the researchers, Pakistani advertising practice has been reported to exhibit more instances of code-mixing; the present corpus, limited to the texts that are English-dominant, reveals a limited amount of code-mixing at the level of lexis, but the patterns of collocation identified are assumed to interact with the Urdu elements in the broader bilingual advertising context. The collocational

structures identified here are simple verb-object structure (get Carfirst), adjective-noun structure (new taste, best offers), and possessive-noun structure (your sleep, your upgrade), which are inherently cross-linguistically transferable and allow seamless code-mixing in the practice of advertising in Pakistan. The comparative study also reveals an important methodological aspect: very few studies in advertising research have used corpus-based collocation analysis in the West, limiting the quantitative cross-cultural comparison. The present work shows that the systematic corpus approach can be used effectively and promotes its use for other advertising corpora in other cultural contexts.

The implications of the above for the discourse of Pakistani advertisement are as follows:

There are a number of important implications of the findings. Lexical conventionalisation is high (as reflected by consistently high MI scores), indicating that billboard writers in Pakistan share a common lexical resource that is conventionalised through repeated use. This conventionalisation has a parallel interpretive impact: the language of city billboards shapes readers' expectations, and once acquired, this set of expectations enables them to comprehend the content of roadside billboards quickly. The tendency of the combinations to cluster around the functional advertising categories (promotional appeal, quality claim, action imperative) indicates that the conventionalisation of this register is functional: clusters become conventional because they perform certain functional and repetitive persuasive functions well.

Urgency and promotional language are key elements of dominance, as this is a competitive retail environment in Pakistan and people are price-sensitive. Insights into the collocational mechanisms of urgency construction (limited offer, available now), value amplification (free spray, best offers) and action messaging (get updated, get digital) are of practical interest for advertising strategy and consumer literacy education. Brand names are integrated into collocational structures, facilitating easier brand recall (e.g., get Carfirst, get telenor, new Vangli), through frequent and predictable brand-linguistic patterns, which has implications for brand awareness strategy. More generally, one needs to ask whether the brand-name collocations created through OAs are transferred to Pakistani English in daily use; this question would be answered through longitudinal corpus research.

Even in the English-dominant corpus, the collocational patterns index the multilingual ecology in Pakistan. Many syntactic and semantic structures are cross-linguistically transferable, allowing natural code-mixing in bilingual advertising contexts. The findings as a whole validate the legitimacy of Pakistani English as a systematic commercial register variety of English in line with the World Englishes perspective of Shah et al. (2019). Last but not least, the successful implementation of LancsBox and MI analysis on this specialist corpus proves the possible extension of the method of corpus linguistics to other registers of Pakistani English (e.g., print media, television commercials, digital advertising, product packaging) to come up

with a more holistic picture of the variation and development of the Pakistani commercial discourse.

6. Conclusion

The findings in this section are summarised in the following section.

This study is the first corpus-based investigation of collocation patterns in Pakistani billboard advertisements, addressing a gap in corpus linguistics, advertising research and World Englishes research. Analysis of the specialised corpus that consisted of 70+ billboard texts (793 tokens) from four major cities of Pakistan (Lahore, Karachi, Islamabad, Multan) using the concepts of LancsBox X 5.5.1 and Mutual Information was carried out on 10 high-frequency node words (MI: 5.5 – 9.6) that showed high conventionalisation of lexical patterns. The major discourse strategies used in Pakistani billboard advertisements are promotional and urgency appeals, quality and improvement claims, and action-oriented imperatives. The dominant discourse strategies in Pakistani billboard advertisements are promotional and urgency appeals (free, limited, offers); quality arguments and improvement claims (new, best, upgrade); and action-oriented imperatives (get, available). The collocational profiles of individual node words shed light on the special ways in which these strategies are realised: free constructs value by means of concrete giveaway associations; limited creates scarcity by means of concrete paired objects; upgrade positions lifestyle enhancement as financially neutral; and your operationalises direct personal address over domains of consumer life.

The presence of brand names (car first, telenor) or national identity markers (Pakistan change, Pakistan consultant) illustrates the localisation of global advertising conventions within the sociolinguistic framework of Pakistan. The study also produces the first collocation frequency list based on MI for advertising English, useful for advertising professionals, English teachers and researchers of World Englishes. The result of this study enriches the empirical foundation of the study of World Englishes by providing a comprehensive picture of the lexical characteristics of Pakistani English in the commercial domain; it also serves as a methodological foundation for further research on South Asian advertising discourse using a corpus-based method. Findings have implications for developing advertising copy in Pakistan using data, designing contextually appropriate English-language teaching materials, and planning global marketing in the South Asian context.

7. Limitations

There are several qualifications that need to be recognised. The corpus size (793 tokens, 70+ billboard texts) is relatively small by general corpus standards, which may lead to higher scores for low-frequency collocations and hamper the generalisability of the results for a broader picture of outdoor advertising in Pakistan. Only 10 node words were analysed from English-dominant texts; Urdu-only and code-mixed patterns are significant parts of Pakistani advertising discourse. The synchronic design of the 2021-2025 design does not allow for diachronic study of language change; if the only data available is MI, additional data, such as log-likelihood and t-score, can be used to gain some other insights; and without multimodal

and audience reception data, the persuasive effectiveness of collocations can be evaluated only on the textual level.

8. Future Research Directions

This study opens several avenues for future research. Increasing the corpus to 10,000+ tokens across different media and time periods will enable more robust analyses and a diachronic study of changes in Pakistani outdoor advertising language. The processing of bilingual texts using multilingual collocation analysis would reveal code-mixing and translingual lexical relationships. Comparisons across nations with advertising corpora from South and Southeast Asian contexts (India, Bangladesh, Singapore, Malaysia) and other World Englishes contexts (Nigeria, Kenya) could detect region-specific versus universal persuasive patterns. The multimodal analysis and audience reception approach, such as eye-tracking research and recall questionnaires, would allow the assessment of the persuasive power of collocations in the eyes of Pakistani consumers. More sophisticated computational methods, such as word embeddings, topic modeling, and machine-learning classification of advertising strategies, might yield collocational dimensions that MI cannot identify. The collocational patterns of the constructions of gender, class, religion and national identity in Pakistani advertisements can be analysed using Critical Discourse Analysis and corpus methods. Another promising direction is pedagogical uses, such as the creation of corpus-based vocabulary materials for Pakistani English learners. Overall, the findings of this study demonstrate the usefulness of corpus-based collocation analysis for analysing the discourse of advertisements in Pakistan and provide a basis for a wider research programme to document the lexical characteristics of Pakistani English in commercial discourse.

References

- Aihua, L. (2013). A corpus-based approach to the linguistic features of advertising English. *Journal of Language Teaching and Research*, 4(1), 75–82.
- Amir, A., Aslam, S., & Ashraf, H. (2025). Examining lexico-semantic diversity in Pakistani English through newspaper corpus analysis. *Annals of Advances in Journalism*, 9(2), 26–38.
- Anwar, B., Shah, S. K., & Sarwat, S. (2025). Lexical variations in the punch lines of marketing the beauty products: A comparative analysis of Pakistani and British English. *Academy of Strategic Management Journal*, 4(4), 943–958.
- Awan, A. G., Azher, M., & Anwar, M. N. (2025). Linguistic innovations in Pakistani English: A corpus assisted study of educational advertisements. *Journal of Applied Linguistics and Translation*, 11(5), 53–64.
- Bartsch, S., & Evert, S. (2014). Towards a Firthian notion of collocation. *Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern*, 48–61.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

- Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22–29.
- Deng, L., Liang, M., & Xu, J. (2022). A multi-dimensional comparison of association measures in collocation extraction. *International Journal of Corpus Linguistics*, 27(4), 451–479.
- Ding, Y. (2009). *Collocations: Experiential grammar and conventionalisation of travel advertising* (Doctoral dissertation). University of Cambridge.
- Evert, S. (2005). *The statistics of word cooccurrences: Word pairs and collocations* (Doctoral dissertation). University of Stuttgart.
- Gugrić, I. (2021). *Adjective-noun collocations in tourist advertising brochures about Istria: A corpus-based translation study*. University of Maribor Press.
- Ha, M. (2023). A corpus-based analysis of semantic differences between “advertising” and “advertisement”. *Studies in Linguistics*, 68, 119–142.
- Habib, H., Khokhar, S. F., & Mustafa, A. (2025). A linguistic analysis of codeswitching in Pakistani billboard advertisements. *Pakistan Social Sciences Review*, 9(2), 51–73.
- Habib, M. (2026). Exploring English-Urdu code-switching in digital advertising: A case study of KFC Pakistan. *Zenodo*. <https://doi.org/10.5281/zenodo.18383888>
- Hafeez, M., Riaz, S., & Haider, S. A. (2025). The linguistic framing and marketing dynamics of digital advertisements: A multidimensional corpus-based study. *Academy of Strategic Management Journal*, 4(3), 560–578.
- Hamid, A., Mahmood, R., & Ameer, S. (2020). Linguistic deviation in Pakistani print media advertisements. *Sustainable Journal of Educational and Social Research*, 3(4), 187–193.
- Imran, M. (2023). The syntactic variations in adverb phrase ‘as well’ in initial position in Pakistani English: A corpus-based study. *World Journal of English Language*, 13(6), 196–205.
- Jablonkai, R. R. (2020). Leveraging professional wordlists for productive vocabulary knowledge. *ESP Today*, 8(1), 4–26.
- Jilani, A. K., Mehmood, M. A., & Iqbal, M. (2018). Lexico-semantic features of Pakistani English newspapers: A corpus-based approach. *International Journal of English Linguistics*, 8(4), 50–59.
- Kolesnikova, O. (2016). Survey of word co-occurrence measures for collocation detection. *Computación y Sistemas*, 20(3), 327–344.
- Lakkam, P., & Manowong, S. (2020). The lexical collocations used in facial skincare product online advertisements. *LEARN Journal*, 13(2), 234–251.
- Lehecka, T. (2015). Collocation and colligation. In P. Durkin (Ed.), *The Oxford handbook of lexicography* (pp. 93–108). Oxford University Press.
- Manan, S. A., David, M. K., & Dumanig, F. P. (2021). English in the linguistic landscape of Pakistan. In *The Handbook of Asian Englishes* (pp. 467–487). Wiley.
- Matte, A. C. F., & Lenz, P. (2019). A quantitative analysis of collocations in Brazilian and British students’ academic writing. *Revista Linguagem & Ensino*, 22(4), 1049–1074.

- Munir, S., Ashraf, H., & Fatima, S. (2025). Linguistic analysis of codeswitching in Pakistani advertisement taglines. *Global Language Review*, 10(1), 78–92.
- Phoocharoensil, S. (2021). Multiword units and synonymy: Interface between collocations, colligations, and semantic prosody. *GEMA Online Journal of Language Studies*, 21(2), 17–34.
- Riaz, S. (2019). Language variation: Code-mixing and code-switching in Pakistani commercials. *Journal of Language and Linguistic Studies*, 15(2), 411–419.
- Sardar, H., Zafar, S., & Iqbal, A. (2020). Lexical variations in Pakistani English: A case study of Coca-Cola and PepsiCo advertising in Pakistan. *Global Language Review*, 5(3), 89–101.
- Shah, S. K., Anwar, B., & Sarwat, S. (2025). Pakistani English and cultural identity: A critical discourse analysis of English-Urdu code-mixing in social media advertisements. *Sustainable Social Sciences*, 4(4), 427–441.
- Shah, S. K., Mahmood, A., & Iqbal, Z. (2019). Critical discourse analysis of Pakistani billboards: A study of language variations in perspectives of World Englishes. *Global Social Sciences Review*, 4(2), 156–165.
- Spicer, N. (2022). Methodological considerations for the use of mutual information. *Research Methods in Applied Linguistics*, 1(2), 100024.
- Stubbs, M. (2009). The search for units of meaning: Sinclair on empirical semantics. *Applied Linguistics*, 30(1), 115–137.
- Sun, Y., Lim, K. H., Kaur, S., Jiang, L., Peng, A., & Maher, S. (2022). Employing natural language processing as artificial intelligence for analysing consumer opinion toward advertisement. *Frontiers in Psychology*, 13, 856663.
- Teodorescu, B. (2015). Linguistic patterns in advertising messages. *Knowledge Horizons – Economics*, 7(2), 126–131.
- Todd, R. W., Khongput, S., & Darasawang, P. (2023). Love is blue: Designing and using a multimodal corpus analysis tool. *Applied Corpus Linguistics*, 3(1), 100042.
- Trung, D. Q., Thu, T. T. M., & Hien, N. L. M. (2025). Persuasive techniques in out-of-home advertising: A comparative study. *International Journal of Business and Social Science Research*, 6(5), 1–15.
- Wermter, J. (2009). Collocation and term extraction using linguistically enhanced statistical methods (Doctoral dissertation). Saarland University.