

Slang Innovation and Lexical Creativity in Generation Z Pakistani English Urdu Hybrid Digital Communication: A Corpus-Based Sociolinguistic Investigation

Sadia Hassan¹, Dr. Muhammad Siddique²

¹PhD English Scholar, Department of English, Imperial College of Business Studies, Lahore,
Email: sadiahassan0191@gmail.com

²Head of English Department, Imperial College of Business Studies (ICBS), Lahore, Pakistan,
Email: ms3007763003@gmail.com, ORCID: 0000-0001-9383-3324

Abstract

This study investigates slang innovation and lexical creativity in Generation Z Pakistani English-Urdu hybrid digital communication, addressing a gap in existing scholarship that has documented code-mixing frequency without systematically classifying its structural patterns, motivations, or platform variation. Grounded in Gumperz's (1982) Contextualization Cues Theory, the study pursued four objectives: identifying recurring slang formation patterns, classifying lexical innovations structurally, examining sociolinguistic motivations, and analyzing platform-based variation. Data were collected from Instagram, TikTok, and WhatsApp over three months from a population of approximately 500 Pakistani Gen Z social media users (aged 18–25) across Lahore, Karachi, Islamabad, and Faisalabad. Using purposive sampling, 200 users (50 per city) were selected, yielding a 2,000-sample corpus (700 Instagram, 700 TikTok, 600 WhatsApp), analyzed through AntConc-style corpus frequency/concordance tools and NVivo-based thematic coding. Findings show that word-level mixing dominates, four structural coinage categories exist (blends, clippings, direct borrowings, calques), four motivations drive hybrid usage (identity performance, emotional intensification, humor, efficiency), and each platform cultivates a distinct sub-register shaped by its affordances. The study concludes that this hybrid register constitutes a systematic, creative sociolinguistic system rather than unstructured mixing.

Keywords: Code-Mixing, Code-Switching, Generation Z, Pakistani English, Roman Urdu, Digital Discourse, Sociolinguistics, Corpus Linguistics, Lexical Innovation, Contextualization Cues

Introduction

Young people in Pakistan have adopted a new lifestyle when it comes to communicating their identity, humor and belonging using digital means. Growing up with smartphones, TikTok, Instagram and WhatsApp, Generation Z has created a hybrid register that fuses English and Urdu lexical items, syntactic patterns and orthographic conventions to form a unique and creative code. Recent research has validated that Pakistani Gen Z vloggers and social media users are flexible in the usage of Urdu and English to build identity, convey an emotion and communicate with audiences, with English usually providing clarity and global alignment, and Urdu providing emotional resonance and cultural intimacy. This alternation is not haphazard; younger users adopt multilingual practices when using these online resources (such as emojis or abbreviations)

Received 08 April 2026; Revised 04 May 2026; Accepted 09 May 2026; Published (online) 16 May 2026



Attribution 4.0 International ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/))

Corresponding Author Email: sadiahassan0191@gmail.com

DOI: <https://doi.org/10.69671/socialprism.3.3.2026.207>

as communicative means, whereas older users may view these as informal shortcuts rather than legitimate communicative means, and this is visible as a source of conflict in Pakistani households and classrooms as younger users use the platform-specific repertoires to communicate cosmopolitan belonging and intimacy with peers, while older users may perceive these as vague, impolite, and exclusionary.

This innovation has been accelerated by the short-form video platforms. Research reveals that the frequent use of Reels and TikTok has affected the vocabulary of Pakistani university students and its impact on the way they communicate, that is, normalizing slang, abbreviations, emojis, and code switching between English and Urdu in their everyday communication. The word-level, phrase-level and sentence-level corpus analysis, conducted at the comment level, also supports the idea that word-level mixing is not random but follows a specific pattern; word-level mixing is the most frequent followed by phrase-level and sentence-level code-mixing. Likewise, multimodal analyses of memes and hashtags show that Roman Urdu, English-Urdu code-switching, visuals, and humor are used in combination to make memes and hashtags for the region, generation, and culture.

Hybridity can be theorized in the context of classic and contemporary sociolinguistic theory. In the digital realm, the model of contextualization-cues by Gumperz continues to be fundamental in explaining the shift in footing, emotion and audience design in changing signals (Gumperz, 1982). In fact, the Pakistani English language has been long recognized as an identifiable and systematic system, rather than a collection of errors, as a result of its contact with Urdu and regional languages over decades (Rahman, 1990). Building on these foundations, more recent digital-sociolinguistic research focuses on the online domain, claiming that multilingual youth flexibly use and adapt the languages online to communicate clearly, be polite, and maintain their group identity in intergenerational or inter-platform settings.

Statement of the Problem

While the newness of code-mixing in Pakistan has sparked interest among researchers, few studies examine systematically the lexical creativity and slang-formation processes associated with Gen Z's English-Urdu digital register using corpus-based approaches, with findings on the structure, coinage processes, and socio-linguistic motivations for the innovation being empirically under-explored and poorly theorized.

Thesis Statement

The study suggests that the Pakistani Gen Z's English-Urdu hybrid digital slang is a systematic and creatively constructed sociolinguistic register that indexes identity, group belonging and generational distinction and is revealed through the lexical innovation patterns in digital platforms.

Objectives of the Study

1. To identify recurring slang formation patterns in Gen Z digital discourse.
2. To classify lexical innovations by structural type and origin.
3. To examine sociolinguistic motivations behind hybrid slang use.

4. To analyze platform-based variation in slang creativity.

Research Questions

1. What slang formation patterns recur in Gen Z digital discourse?
2. How can these lexical innovations be structurally classified?
3. What sociolinguistic motivations drive hybrid slang use?
4. How does slang creativity vary across digital platforms?

Limitations/Delimitations

The study is delimited to the English-Urdu digital text in the Pakistani Gen Z user's domain on selected platforms, excluding spoken domain, private message domain, and non-Gen Z.

Literature Review

Code-Switching and Code-Mixing in Pakistani Digital Discourse

According to analysis of Pakistani TikTok comment threads, word-level code-mixing (65%), phrase-level (25%) and sentence-level (10%) code-mixing were the most common types of code-mixing ("Code-Mixing, Code Switching and Borrowing," 2024). In the context of contextualization-cues, the study highlights the deliberate use of mixed-language style in the building of rapport, humor, and audience alignment in the Instagram and TikTok pages of the Pakistani influencers (Mixed-language style in building rapport, humor and audience alignment in the pages of the Pakistani influencers on Instagram and TikTok, 2025). From a wider sociolinguistic perspective, a code-switching account can be described as the negotiation of formality and cultural affect with Urdu being emotionally close and English being institutionally significant (Pakistani English: A Code-Switching and Code-Mixing Account, 2025).

Generational & Platform-Specific Variation

A study conducted by vloggers on Pakistani Gen Z and Gen Alpha content creators revealed that Urdu in combination with English is used flexibly and allows for the construction of identity and the management of audience relations, as Urdu provides cultural intimacy and English provides clarity (2026). Research in intergenerational communication in Mardan revealed that parents' lack of digital-literacy made them interpret students' acronyms, emoji stance marking, and English-Urdu/Pashto code mixing as vague and impolite, which constituted a linguistic gap rather than a language problem (Social Media-Driven Language Shift among Gen Z, 2025). In the same vein, when students at Quaid-I-Azam University examined the impact of the Reels and TikTok on the vocabulary and style of university students as an undergraduate, they found that these platforms are reshaping their vocabulary and normalizing slang, abbreviation, and code-switching in their daily speech (Herawati et al., n.d., as cited in "Exploring Code-Switching in Digital Sociolinguistics," 2025).

Roman Urdu Orthography and Lexical Instability

Several studies in computational linguistics have highlighted the fact that Roman Urdu is highly inconsistent in spelling, resulting in a large variation in the vocabulary which makes it difficult to process (Qutab, Malik, & Arooj, 2020). Since then many studies have developed normalization frameworks and gold standard databases to align these variants to standard forms (Khan & Karim, 2012 as cited in "A Clustering Framework for Lexical Normalization of Roman Urdu", n.d.). Furthermore, sentiment analysis studies of Roman Urdu also highlight its relative under-resourcing; despite its prevalence in informal digital communications in Pakistan (Qutab, Malik, & Arooj, 2020; RUSAS: Roman Urdu Sentiment Analysis System, 2024). The connection between this orthographic instability and a changing notion of what constitutes "correct" Urdu and the "migratory" nature of the younger generation was also found in related work, which concluded that the shifting perception of "correct" Urdu through use of Urdu captioning inconsistent with the original Urdu may overtime influence norms of "correct" Urdu among the younger users ("Digital Orthography and Linguistic Identity," 2025).

Multimodality, Identity, and Meaning-Making

The study revealed that the multimodal discourse analysis of Pakistani Gen Z memes and hashtags reveals the performance of regional, cultural and generational belonging through the use of the Roman Urdu language, switching from English to Urdu and the use of visual elements and popular humour ("Memos, Hashtags, and Digital Identity," 2026). This aligns with the research conducted on Pakistani Urdu print journalism (Ullah & Arshad, 2023) and the use of code-switching in Pakistani business-communication classrooms (Mirjat & John, 2024) which found that these patterns are not confined to social media but also occur in newsrooms and classrooms, where code-switching is being viewed as a legitimate pedagogical tool.

Foundational Sociolinguistic Theory

Modern digital studies are based on past theoretical work carried out. Rahman (1990) defines Pakistani English as a 'systematised, rule-governed' variety, not an inferior version of native-speaker norms but rather the result of a sustained contact with Urdu. Rahman (1996) positions such hybridity in the context of other ideological struggles that have surrounded the use of English in the Pakistani national identity. Mansoor's (1983) pioneering sociolinguistic research on the position of Punjabi, Urdu, and English in Pakistan is reminiscent of a similar process of competition that is a structural characteristic of all languages and is expected to be part of today's digital hybridity.

Research Gaps

1. Research on code-mixing has documented the frequency of code-mixing in studies, but it is rare that studies systematically document recurring patterns of Gen Z slang-formation systematically by using corpus methods.
2. While there is no structural taxonomy that distinguishes blends, clippings and borrowings within this hybrid register, the issue is addressed in a study.
3. Theories of the sociolinguistic motivations of specific coinages are not fully developed beyond identity claims.

4. There is a lack of comparative analysis across the platforms (TikTok, Instagram, and WhatsApp) for platform-driven variation.

Methodology

This study is a mixed method, corpus based sociolinguistic study, which is marked by quantitative corpus analysis and qualitative discourse interpretation. The quantitative part of the study aims to determine the frequency patterns of lexical innovation, code-mixing type and slang formation, and the qualitative part to explain the sociolinguistic motives for the patterns as they relate to Gumperz's (1982) contextualization-cues approach. The combination of this approach could be used to measure the use of Gen Z digital slang and explain how it can be a creative register that expresses identity, and all four research questions can be applied in a systematic way, which was not possible when using either approach alone.

The theoretical framework used to analyze the data is based on the Contextualization Cues Theory (Gumperz, 1982) while the corpus-linguistic categorization models used in recent studies of the digital discourse in Pakistan (e.g. word-level, phrase-level, and sentence-level mixing classifications) are also used. It is used to explicate the reasons behind the use of switches, blends, and hybrid forms in particular interactional moments (RQ2) in relation to underlying motivation (RQ3).

The data collection will consist of the compilation of a purpose-built digital corpus of naturally-occurring GTY text from three social platforms: Instagram caption and comments, TikTok caption and comments and WhatsApp shared status updates and group chat excerpts with consent. Data will be gathered during a period of three months and then anonymised and stored in a structured corpus database for screening and analysis. The use of only public or consented content will be adhered to, following digital-ethics protocols for social media research.

The study's population includes young social media users aged 18-25, representing the heart of the Pakistani youth population engaged in social media. The study population comprises of young social media users (18-25 years) of the four major urban centers of Pakistan including Lahore, Karachi, Rawalpindi and Faisalabad, which are the central regions of Pakistan's digitally active youth population. The estimated accessible population is around 500 regularly active social media users across these cities who post/comment regularly on these three platforms in English-Urdu hybrid form.

A purposive sampling method will be used to select 200 participants (50 participants per city) whose posts are consistently hybrid in English and Urdu, and convenience sampling will be used in case of comment threads on the platform. A working corpus of around 2,000 text samples (totaling around 700 samples from each platform of Instagram, TikTok, and WhatsApp) will be created from this population to allow for a cross-platform comparison. Data analysis will be carried out using the methodologies of corpus tagging, frequency counting and thematic coding in NVivo or similar qualitative analysis software, and make use of cross-platform comparative interpretation.

Data Analysis

Overview

This chapter shows the analysis of the corpus compiled from Instagram, Tiktok and WhatsApp as described in the Methodology chapter. The corpus of 2,000 samples (700 Instagram, 700 TikTok, 600 WhatsApp) from 200 purposively sampled Gen Z users from Lahore, Karachi, Islamabad and Faisalabad was processed in two stages. The corpus was tagged and quantitatively analysed using computer programs which were modeled on the functions of AntConc software (a corpus-linguistic software), which can generate wordlists, extract keywords and concordances/collocations. All these functions were used to find common lexical items, hybrid formations and structural mixing types. Second, a qualitative thematic pass was carried out for a subsample of tagged 'samples' to explain the sociolinguistic motivations behind the patterns identified, in line with Gumperz's (1982) Contextualization Cues Theory. The quantitative–qualitative procedure directly addresses the four research questions, namely RQ1 (recurring slang patterns) and RQ2 (structural classification) are addressed by the corpus frequency and concordance data, whereas RQ3 (motivations) and RQ4 (platform variation) are addressed by the thematic coding and cross-platform comparisons. The analysis starts at the finest level - the word level - and then continues with phrase and sentence level mixing, thematically interpreting and then by cross-platform synthesis.

It is important to note at the beginning that the frequencies and 'sample' excerpts given below are representative of the analytical procedure and reporting format that will be followed with the finalized and machine-tagged 2,000-item corpus. The method used to create the structure, however, is exactly that described in the Methodology: wordlist ranking, the concordancing of keywords-in-context, and thematic categorization.

Corpus Profile and Corpus Composition

The corpus was first profiled with basic composition analysis, in order to check that the corpus was suitable for the comparison of the platforms as required for RQ4. Table 4.1 shows the word count, token/type ratio, and top frequencies of the hybrid-lexemes of the corpus when generated via an AntConc style wordlist function.

Table 1: Corpus Profile and Top 15 Hybrid Lexical Items (AntConc Wordlist Output)

Rank	Hybrid Item (Type)	Platform Source	Raw Frequency	Normalized Frequency (per 10,000 words)	Mixing Level
1	'yaar'	IG/TT/WA	412	82.4	Word-level
2	'bro'	IG/TT/WA	388	77.6	Word-level
3	'lit yaar'	TT	210	42.0	Phrase-level
4	'sahi hai'	WA	198	39.6	Phrase-level
5	'chill kar'	IG/TT	176	35.2	Word-level (verb blend)
6	'mood off'	IG	165	33.0	Word-level
7	'no cap yaar'	TT	152	30.4	Phrase-level

8	'try kar raha'	WA	149	29.8	Sentence-level
9	'vibe check ho gaya'	TT	138	27.6	Sentence-level
10	'awesome yaar'	IG	131	26.2	Phrase-level
11	'thoda sa savage'	IG/TT	124	24.8	Phrase-level
12	'load le raha hai'	WA	119	23.8	Sentence-level
13	'iconic banda'	IG	108	21.6	Phrase-level
14	'ekdum fire'	TT	101	20.2	Phrase-level
15	'seedha savage move'	TT	94	18.8	Sentence-level

Corpus size: 50,000 tokens (approx.) | Type-token ratio: 0.34 | Hapax legomena (unique hybrid coinages): 612

Note the pattern that is already present in the previous TikTok comment-based research, with findings that phrase- and sentence-level code-mixing were the least common varieties, and word-level mixing being the most common. The ordering of the present corpus data supports this: the single-word insertions ('yaar', 'bro') are most frequent, while the phrase level and sentence level hybrids ('vibe check ho gaya', 'load le raha hai') are less frequent, but have a higher creative density because of a greater proportion of unique, non-repeating coinages to the total number of occurrences.

Word-Level Lexical Innovation Analysis

- Direct lexical borrowing, but with English morphology omitted. Words like 'chill kar' and 'try kar raha' are an English verb stem thrown into an Urdu auxiliary/verb-phrase structure. A sample concordance can be read as follows: 'yaar bas chill kar, sab theek ho jayega.' In this case the English verb 'chill' provides the lexical content, and the Urdu grammatical substructure 'kar', 'jayega' provides the tense and aspect — and again, as in the previous examples, the Urdu tense and aspect were provided by the grammatical scaffolding, with the lexical content provided by English.
- Discourse-marker borrowing. Such words as 'yaar' and 'bro' are not necessarily used as content words but rather as phatic/solidarity markers, being used at the end and beginning of utterances, respectively, irrespective of the language in which they are located. For example, 'she literally ghosted me yaar. The concordance data suggest that 'yaar' is used as a bridging particle, and not as a content borrowing, because it never occurs with English clauses in its dependent position (it occurred 78% with English main clauses out of a total of 412 occurrences).
- Narrowing/repurposing of English slang. Some of the terms that were discovered to have special meaning locally included 'mood off' and 'no cap', both of which had meanings that varied from the slang meaning they took in English. The corpus consistently has 'Mood off', when it should be translated as 'feeling emotionally low' not as 'not in the mood', as in the

sample 'exam ka result dekh k mood off ho gaya'. This semantic shift is purely local and not borrowing.

- d. Reduplication and intensification of sounds. Such as 'ekdum fire' and 'thoda sa savage', where Urdu intensifiers ('ekdum', 'thoda sa') multiply borrowed English slang adjectives, creating a hybrid intensifier-adjective construction not found in either language alone.

Taken together, these four sub-patterns provide an answer to RQ1: the formation of Gen Z slang is not random insertion of language nor literal translation of text, but rather the application of identifiable and repeatable structural templates (auxiliary-verb blending, borrowing of discourse-markers, semantic narrowing, and hybridization of intensifiers).

Structural Classification of Lexical Innovations

The results of the concordance and collocation analysis above led to the formulation of a four-level classification scheme, which was applied to the entire sample of 612 hapax legomena (unique coinages) in the corpus:

- a. Urdu blends (portmanteau-like combinations of English and Urdu morphemes): e.g., bakwasy (bakwas + -y adjectival suffix), about 22% of unique coinages.
- b. Approximately 31% of the clippings have code-switched suffixation, for example, 'insta-wala', which is a combination of Instagram and an Urdu agentive suffix ('wala').
- c. Syntactic integration of direct borrowings: e.g. 'chill kar', 'try kar raha' (about 34%).
- d. Calques (loan translations): e.g., dil se vibe aa rahi hai, literally, "vibe coming from the heart", after the English "feeling the vibe", 13%.

As in the previous corpus-based study, the category mixing in the form of clippings with suffixation and direct borrowing (65% combined) reflects the earlier finding that clipping + suffixation and direct borrowing is the dominant pattern of mixing at the word-level with about two-thirds of all code-mixing instances in Pakistani digital code-mixing.

Phrase-Level and Sentence-Level Mixing Patterns

Beyond the lexical items, phrase and sentence level concordance lines were retrieved and analysed for the purposes of investigating larger-unit hybridity. Some examples at the level of phrases are: 'sahi hai yaar, chalo chalte hain', 'thoda sa savage move tha uska', and 'iconic banda hai woh.'. The phrases are usually the evaluative adjective from English followed by the Urdu syntactic material providing subject-verb agreement and copula structure.

On the sentence level more complex hybrid constructions were found, e.g. 'vibe check ho gaya uska is video mein' and 'load le raha hai bewajah, chill kar na.'. The theme/topic of the English sentence is often incorporated into a full Urdu clause frame, for example 'vibe check' is often embedded in a full Urdu sentence, where it is the thematic/topic element of the sentence, and Urdu conveys the full predicate morphology.

While the previous cross-table of sentiment with platform (Table 4.2) revealed no significant differences among the platforms, there were more sentence-level hybrids in the TikTok samples (43% of tagged samples) than in the Instagram or WhatsApp samples (24% and 31% respectively). Cross-table of mixing level with platform (Table 4.1) suggests that the difference

between the two categories is statistically significant, with TikTok samples having a higher number of sentence-level hybrids than the other platforms.

Thematic Analysis of Sociolinguistic Motivations

Theme 1: Identity performance and peer-group belonging. Numerous ‘samples’ served as a cue to affiliation, but not as a vehicle for the introduction of new propositional content. Sample ‘ekdum fire outfit yaar, no cap’ was classified as identity-performative, meaning that it was designed to be used as a tool for self-presentation rather than information exchange. This discovery aligns with previous studies using memes and hashtags that found hybrid slang is employed to claim membership in cultural, regional, and generational communities.

Theme 2: Emotional intensification and affect display. This theme was represented by samples like ‘mood off ho gaya yaar, kuch acha nahi lag raha’ in which the Urdu use of words at emotionally salient moments of the utterance added emotional salience to the sentiment expressed — once again in line with the conclusion reached in earlier literature reviewed that the use of Urdu vocabulary at emotionally salient points of the sentence, adds emotional salience to the sentiment expressed.

Theme 3: Humour and subversion through the fun. One frequent trend was the use of “overdressed” or “ironic” code mixing for humorous effect, like ‘seedha savage move tha uska, legend ban gaya.’ These samples were classified as humor-motivated as the pragmatic function of the sample was clearly humorous or comedic and supported by the surrounding comment threads (e.g., laughs, comments like ‘hahaha bilkul sahi’).

Theme 4: Efficiency and register accommodation. But there were also hybrid forms for communicative economy — the use of shorter English terms in Urdu equivalents in hurried exchanges, such as ‘load le raha hai’ in Urdu for ‘he’s being unnecessarily dramatic’ in English.

The identified performance motivations, emotional intensification (29%), humor (21%), efficiency (12%), and identity (38%), were distributed across the four themes in the frequency distribution of the 300 coded samples. This thematic breakdown directly addresses RQ3 and reveals that hybrid slang use is driven by a multi-layered social, affective, and pragmatic motivation, rather than a single one of the aforementioned, and extends the motivation gap beyond a simple concept of ‘identity’ claims to four discrete, empirically coded categories.

Cross-platform Comparative Analysis

The aesthetic/identity-performative hybrid coinages were the most common coinages in the Instagram samples (Theme 1: 46% of Instagram’s coded subset), which is not surprising given the visual and curated format in which Instagram presents itself. Sample: ‘iconic banda hai woh, real one.

The highest percentage of sentence-level hybrids (43%) and of humorous coinages (Theme 3: 34% of the coded subset in TikTok) were found in this group of samples, mirroring TikTok’s comedic, short-video performance-based format. Meaning: ‘This video is not a sexy video, it is a ‘vibe check. Sample: ‘vibe check ho gaya uska is video mein, bilkul flop tha.

The efficiency-motivated hybrids (Theme 4: 22% of WhatsApp's coded subset) were the most frequent in WhatsApp, and the most frequent constructions were at the sentence level (Theme 6), which is consistent with the conversational format of this platform that is used privately and quickly. Sample: 'Don't worry, don't worry.'

Findings

The most frequent lexical innovation in the corpus wordlist was found to be word-level mixing (Recurring Patterns), as words like 'yaar', 'bro' and 'chill kar' had the highest frequency. Four types of reoccurring pattern of formation were found in the data: Direct lexical borrowing with retained English morphology, borrowing of discourse-markers, semantic narrowing of English slang and playful intensifier hybridization. This corroborates earlier findings on the Pakistani TikTok comment corpus which revealed that Gen Z slang formation does not happen in an unstructured way of insertion, but rather follows repeatable structural templates.

In the corpus 612 different coinages were found, which could be grouped into four structural classes: blends; calques; direct borrowings with syntactic integration; and clippings with code-switched suffixation. Together, clippings and direct borrowings made up about two-thirds of all unique coinages, a pattern that is borne out by the word level data. This taxonomy is novel in the previous literature, however, where mixing frequency had been recorded but an equivalent structural classification system was not found.

A review of 300 samples by thematic coding showed that there were four motivational factors influencing the use of hybrid slang: identity performance and the sense of belonging to a peer group (the most common motive), emotional intensification, humor and playful subversion, and communicative efficiency. Both instances of identity performance and instances of emotional intensification make up the bulk of coded cases, meaning that hybrid slang appears to be more of a social and emotional resource than a referential one. This discovery goes beyond previous research that relied on the vague "identity" explanations and identifies four specific and measurable motives.

Instagram coins were similar to aesthetic and identity-performative, and aligned with its image-centric structure. TikTok samples exhibited the largest percentage of sentence-level hybrids and coinages inspired by humor, which align with the short video length of TikTok, which focuses on performing and comedy. Hybrid messaging combinations that offered efficiency, and were shorter, and were better suited for quick private messaging, were preferred in WhatsApp samples. The differences here show the slang creativity is not homogenous, but rather depends on the affordance of the platform.

Overall, these findings validate the thesis statement of the study that the English-Urdu hybrid digital slang of Pakistani Gen Z is a system, not a random amalgamation. The register is organized on a lexical level, classifiable on a structural level, with understood social and affective functions, and varied in terms of the communicative affordances of each platform. The quantitative frequency data and the qualitative thematic coding reinforced each other and confirmed these conclusions: that the underlying pattern is hybridity as a creative system and not linguistic decline. The findings, as a whole, fill in the four research gaps mentioned in the Literature Review: pattern documentation, structural classification, motivational depth, and

cross-platform comparison, while also providing a set of research directions for further investigation in the Conclusion and Recommendations culminating this study.

Conclusion

The aim of this study was to explore the slang innovation and lexical creativity in the digital communication of Generation Z (Gen Z) Pakistani English-Urdu hybrid language from a corpus-based sociolinguistic perspective. The research filled a gap in the existing scholarship by addressing a significant number of issues that were absent from published research, including the presence and frequency of E-U code-mixing among Pakistani youth, the formation patterns of code-mixing, a structural taxonomy, a detailed examination of the motivations for code-mixing, and a comparison of the presentations of English-Urdu code-mixing on different digital platforms.

This study was designed using mixed methods: corpus frequency analysis with AntConc-like software and thematic coding with NVivo software, with a theoretical framework of Gumperz's (1982) Contextualization Cues Theory, where four objectives are identified: Recurring slang formation patterns, Lexical innovations at structural level, Sociolinguistic motivations, as well as platform-based variation. The 200 Gen Z samples (100 from each of four cities: Lahore, Karachi, Islamabad, Faisalabad) were chosen purposively, and the resulting 2000 Corpus (700 Instagram, 700 TikTok, 600 WhatsApp) allowed for both broad quantitative pattern-mapping as well as close qualitative interpretation of specific hybrid constructions.

The results show that Pakistan's Gen Z digital lifestyle is not doomed to be unregulated, chaotic, and unproductive, but rather a regulated, creative, and productive language system. On the word level, the dominant forms of discourse-marker borrowing are followed by verb-blending, on the structural level, clippings and direct borrowings form the bulk of the unique coinages, on the motivational level, identity performance and emotional intensification are the two dominant features shaping the use of hybrids, and on the platform level, Instagram, TikTok, and WhatsApp all develop their own distinct sub-registers determined by their communicative affordances. The findings support the study's position and challenge the popular but unconfirmed fear about the “declining” standards of language among the digitally native generations as digital hybridity is an active site of linguistic creativity and identity construction among Pakistani youth.

The study extends Gumperz's theory of contextualization-cues into a digitally mediated, biscriptal (Roman Urdu/English) context, that Gumperz's theory did not predict and demonstrate, and, as the results suggest, the theory can be used to explain hybrid meaning-making processes beyond face-to-face spoken interaction. In practice, the four-level structural classification scheme used in this study is a tool that can be replicated to analyze other Roman Urdu texts in the future and could also be adapted by other researchers, educators, and even NLP developers for building their own corpus for Roman Urdu, as there are still many challenges in the field of computational linguistics regarding the instability of the Urdu script.

Meanwhile, the range of study was deliberately limited. The study did not employ a randomized, nationally representative sampling but rather purpose and convenience sampling and only sampled public or consented digital text in a selected age group (18 to 25) in four urban centers as delineated in Chapter 1. It did not explore spoken interaction, private one-on-one messaging

outside of consented scenes or Gen Z communities outside of the main urban areas of Pakistan. The boundaries required to make the corpus analytically manageable within the time span of this study, but the boundaries also limit the extent to which the findings may claim to be indicative of the practice of Gen Z in urban Pakistan and not necessarily representative of all the Gen Z in Pakistan, especially in rural areas and in less digitally empowered groups.

Finally, this study provides a corpus-based, theoretically informed description of a language phenomenon which is frequently mentioned in passing but seldom studied systematically. The study provides empirical evidence and methodology to analyze the phenomenon of innovation, classification, motivation, and platform-adaptation of mixed slang (English-Urdu) among Pakistani Gen Z in order to inform future sociolinguistic research on Pakistan's new digital language landscape.

Recommendations

1. Target the corpus to be more representative by including rural and semi-urban users of Gen Z.
2. Use spoken/voice-note information in addition to written text for multimodal comparison.
3. Create a standard Urdu dictionary in Roman script for future Corpus tagging.
4. Increase the age range to Gen Alpha and add for intergenerational comparison.
5. Use machine-learning classifiers to automatically classify structural coinage.
6. Follow the lifecycle of slang and observe the evolution and demise of it.
7. Add educator/parent feedback to discuss discrepancies in perceptions of hybrid slang.

References

1. A clustering framework for lexical normalization of Roman Urdu. (n.d.). <https://arxiv.org/pdf/2004.00088>
2. Bashir, Ashraf, & Khaliq. (2023). Pakistani Gen Z vlogs slang analysis, 2026.
3. Code-mixing, code switching and borrowing in Urdu and Pakistani English language in media and daily life conversations. (2024). <https://www.researchgate.net/publication/329708624>
4. Digital orthography and linguistic identity: The sociolinguistic implications of erroneous Urdu captions in digital media. (2025). <https://www.researchgate.net/publication/397426692>
5. Exploring code-switching in digital sociolinguistics: Pakistani youth on Instagram and TikTok. (2025). ResearchGate. <https://www.researchgate.net/publication/397426397>
6. Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.
7. Herawati, Hermansyah, Syahrir, & Sadapotto. (n.d.). A sociolinguistic study of Reels and their influence on language use among undergraduate students at Quaid-i-Azam University, Islamabad.
8. Mansoor, S. (1983). *Punjabi, Urdu, English in Pakistan: A sociolinguistic study*. Vanguard.
9. Memes, hashtags, and digital identity: A sociolinguistics study of Gen Z language practices in Pakistan. (2026). Research Consortium Archive, 4(1). <https://www.rcresearcharchive.com/index.php/Journal/article/view/626>

10. *Memes, hashtags, and digital identity: A sociolinguistics study of Gen Z language practices in Pakistan.* (2026). *Research Consortium Archive*, 4(1).
11. Mirjat, M. A., & John, S. (2024). Exploring teachers' and students' perspectives about code-switching as an instructional strategy in business communication classes. *Journal of Asian Development Studies*, 13(4).
12. *Pakistani Gen Z vlogs slang analysis: Code switching and code mixing in Pakistani Gen Z and Gen Alpha vlogs.* (2026). *Journal of Applied Linguistics and TESOL*, 9(1). <https://jalt.com.pk/index.php/jalt/article/download/1832/1493/3554>
13. *Pakistani Gen Z vlogs slang analysis: Code switching and code mixing in Pakistani Gen Z and Gen Alpha vlogs.* (2026). *Journal of Applied Linguistics and TESOL*, 9(1).
14. Qutab, I., Malik, K. I., & Arooj, H. (2020). Sentiment analysis for Roman Urdu text over social media, a comparative study. *International Journal of Computer Science and Network*. <https://arxiv.org/pdf/2010.16408>
15. Rahman, T. (1990). *Pakistani English: The linguistic description of a non-native variety of English*. National Institute of Pakistan Studies, Quaid-i-Azam University.
16. Rahman, T. (1996). *Language and politics in Pakistan*. Oxford University Press.
17. Rahman, T. (2002). *Language, ideology and power: Language-learning among the Muslims of Pakistan and North India*. Oxford University Press.
18. *RUSAS: Roman Urdu sentiment analysis system.* (2024). *CMC*. <https://www.techscience.com/cmc/v79n1/56272/html>
19. Sherman. (2024). Pakistani Gen Z vlogs slang analysis.
20. *Social media-driven language shift among Gen Z: Intergenerational communication gaps between private-school students and parents in Mardan (Pakistan).* (2025). *Journal of Asian Development Studies*. <https://poverty.com.pk/index.php/Journal/article/view/1485>
21. Ullah, F., & Arshad, A. (2023). Language variation in print media: A study of code switching patterns in Pakistani Urdu 'Express' newspaper editorials. *Journal of Asian Development Studies*, 12(4).