

DIGITAL VERNACULARS IN PAKISTAN: A CORPUS-BASED STUDY OF LEXICAL INNOVATION AND CODE-MIXING ON INSTAGRAM

Sumaira Ghaffoor

M.Phil. Scholar, Department of English, University of Okara, Okara, Pakistan

Email: sumairaghaffoor345@gmail.com

Ahtsham Ilahee

M.Phil. Scholar, Department of English, University of Okara, Okara, Pakistan

Email: ahtshamilahee143@gmail.com

Dr. Muhammad Ilyas Mahmood (Corresponding Author)

Assistant Professor, Department of English, University of Okara, Okara, Pakistan

Email: ilyas.edu.tesl@gmail.com

Abstract

This paper assesses changing tendencies in digital linguistic activities in Pakistan based on a corpus analysis of Instagram posts with special consideration of lexical innovation and code-mixing. Based on the dataset of 1,200 public Instagram posts by Pakistani users in 2024-2025, the study shows new trends in the use of hybrid linguistic expression depending on globalization, technological progress, and local sociolinguistic principles. It was revealed that the English, Urdu, and regional languages have dynamic relations under digitally mediated situations, which demonstrates the creative ways of code-mixing, transliteration, and blending the style by the users. These practices do not just demonstrate identity performance but also allow expressing emotions and communicating with the various audience. Using sociolinguistic theories and analysis of digital discourse, this paper puts Instagram in a new perspective as a space to share images, but it turns out to be a place where identities are negotiated, and languages are innovated.

Keywords: *Code-mixing, Instagram, Pakistani youth, digital communication, lexical innovation, transliteration, sociolinguistics, identity, globalization and language practices*

1. Introduction

The growing number of contextually-sensitive neologisms, performative forms of bilingualism, and hash tagged code-mixing is a relevant phenomenon that can signify the changing linguistic behavior of Pakistani young people in digital vernaculars. The research contributes to the conceptual and methodological understanding of the corpus of multilingual online discourse, thus enhancing the wider scholarly debate on the topic of digital sociolinguistics, language ideology and media literacy in the post-colonial setting.

The Internet is becoming a connective artery that connects people on different continents, which leads to sudden changes in the linguistic practices in the modern digital society. Platform affordances, socio-political contingencies, and local norms of communicative practice mediate the evolution. These changes are particularly vivid within the Pakistani framework, a multilingual community with a complicated structure of language ideologies, which is particularly evident on such platforms as Instagram. This context, the young generation that can be described as the digital vanguard of Pakistan, engages in a subtle language negotiation process, which combines personal and group identity. Pakistani young people are actively Tamilizing the traditional boundaries of language through ingenious code-mixing, the use of transliterated neologisms, stylistic experimentation, and other means, thus establishing a digital lexicon that cuts across linguistic, scriptural, and linguistic registers, challenging the linguistic orthodoxy.

The virtual landscape of Pakistan is becoming more and more determined by the swift growth of the Internet penetration, which currently is over 129 million users (Pakistan Telecommunication Authority, 2025). More specifically, the young audience segment is quite essential, as it forms the Centre of this emerging virtual environment. Recent data show that about 64 percent of the Pakistani people are less than 30 years old; a significant share of them participates in the social media, with Instagram becoming one of the most successful platforms to communicate digitally (World Bank, 2024).

In Pakistan, digital ecosystem is a promising source of language innovation and language hybridism, as Blommaert and Varis (2015) have addressed it, defining these phenomena as polycentric vernaculars. The characteristic of these vernaculars is that they are neither formal, nor rigid like standard language varieties and neither can be controlled by a central authority; instead their stylistic difference is expressed in many loci of power.

Such polycentric types of communication thrive in the active digital environment, integrating text, visual, and interactive messages. Instagram in Pakistan can serve as a social site, not only as a means of communication, but as a place of linguistic performance as a means of identity formation and negotiation by younger users. The digital space is defined by a permeation of English, Urdu, and other local languages that is usually expressed in unpredictable, multimodal speech. The affordances of Instagram, such as hashtags, comments, stories, and visual images, enable users to express themselves in a variety of ways, using both linguistic and visual signs and symbols, e.g. emojis, to create complex communicative practices. As a result, the differences between formal and informal languages, global and local languages, and various scripts are further diluted. The hybrid digital lexicon witnessed on Pakistani Instagram is not only a form of an informal lexicon, but also a form of linguistic agency of the younger generation who is creative in synthesizing the components of different linguistic systems and fuses them into forms that have a social meaning.

Examples of this practice would include expressions like: mood off hai, a lexicon, a combination of English and Urdu; pindi swag, a mix of a local Urdu word and an English word; and unfollow hogaya, an amalgamation of English and Urdu common to local youth street language.

These expressions cannot be comprehended as simple linguistic shortcuts; instead, they have a deep social and cultural meaning. They are indicative of lingual activities of Pakistani youths, whereby members cue group membership, bargain identity and define their places within both local and global networks. As a phenomenon of identity and status that is emerging on the digital front, these examples of identity and status can inform the modern social dynamics of Pakistani youth culture.

Pennycook (2007) says that a language is not a system but a social practice which is part of daily life.

Language plays the role of an identity claim, affiliation building, and a sense of belonging in specific online communities on the Instagram online platform. In this respect, Instagram represents a kind of spatial locus where language is changing and evolving constantly, reflecting the changing tendencies in youth culture and the dynamics of online communication. To the Pakistani youth, language and script mixing is a creative act that reflects the experience of living in a world that is globalized, local, and cultural practices are merged. Although much of the research has focused on bilingualism, code-switching, and multilingualism in Pakistani classrooms (Mansoor, 2005; Awan et al., 2025) and among urban young people (Butt et al.,

2025) in general, there is still a general lack of literature discussing the appearance of these phenomena in the quasi-public and visual-oriented environment of social media applications and Instagram specifically.

The current paper aims to fill this gap by exploring the use of code-mixing and language innovation by young Pakistani Instagram users. Using the mixed-methods paradigm, the study gathers a specific corpus of the Instagram material produced by Pakistani users during the timeframe of 2024-2025 to, thus, conduct an analysis of how these linguistic practices evolve in the digital realm. The research paper through this empirical investigation leads to the generalization of the digital language practices in the post-colonial societies with special reference to how Pakistani youngsters are setting new terms to language in the environment of modern globalization.

This research, therefore, adds to the growing body of knowledge on digital sociolinguistics and provides useful information about how new media affects the language evolution, identity construction, and communicative processes in the Pakistani society with a multilingual setting. The results explain how digital platforms, such as Instagram, serve as spaces where linguistic identities are negotiated and where new forms of cultural expression by the younger generation of Pakistan emerge.

2. Literature Review

With the modern digital era, social media networks have become one of the key elements that has defined the communication culture especially among the youth segment. The various practices of digital communication like on Instagram have experienced a dramatic linguistic change in Pakistan which is a society where multilingualism is a hallmark. The fact that the languages mix in the online space, this means code-mixing, transliteration, and a range of stylistic inventions, shows the influences of global and local levels that define the communicative practices of the Pakistani youth. The current literature analysis considers the changing language activities on Instagram in Pakistan, and the concept of code-mixing, identity negotiation and hybridization of language forms, which are all relevant in the context of the digital environment.

2.1 Language Practices in Pakistan's Digital Space

Pakistan is a multilingual country that more than seventy languages are spoken, and Urdu and English are the national and official language, respectively (Jabeen and Memon, 2019). This linguistic diversity can also be applied to the digital space in the country because the Internet can be perceived as a living form of linguistic creativity. As stated by the Pakistani Telecommunication Authority (2025) Pakistan has over 129 million internet users and the youthful population forms a huge percentage of this online consumer base. This is a group of people who use social networking sites like Instagram to a great extent, and in this platform, they bargain identities, form social meaning, and exchange with global and local cultural elements using language.

In this respect, social networking sites like Instagram serve as the site of sociocultural exchange and linguistic performance. The notion of polycentric vernaculars is proposed by Blommaert and Varis (2015), who refer to the informal, unstable forms of language that arise in a number of different normative centres, as opposed to being controlled by one authority. Instagram in Pakistan is the example of this polycentric vernacularity, as the English, Urdu,

and regional languages, such as Punjabi, Pashto, and Sindhi, mix. Participants in this platform obtain actively involved in a digital hybridization of language, a combination of registers and scripts in an approach that, however, goes beyond other traditional linguistic limits.

2.2 Code-Mixing and Linguistic Hybridization

Code-switching, which refers to the use of two (or more) languages in a particular communication process, is a thoroughly-documented linguistic process in Pakistan (Mansoor, 2005). It is most widespread in the Internet, where it serves as an identity-making tool, a sense-making tool, as well as a tool of affective signalling. The users of Instagram constantly interchange English, Urdu, and regional languages in single posts or even comments, thus showing linguistic elasticity and accommodative adaptation to various social environments. As an example, such phrases as mood off hai are a mixture of English and Urdu words, whereas such expression as pindi swag is an amalgamation of local and international cultural code. These lexical blends are not simply informal slangs; these are in fact more a process of negotiating and innovating linguistically and young speakers can negotiate between local and global cultural codes.

Code-switching in the Pakistani digital world is a dynamic identity-making activity. Gumperz (1982) states that code-switching and code-mixing are inextricably connected to social identity of interlocutors who have to negotiate between two or more linguistic and cultural regimes. The use of social media, especially Instagram, provides a setting as Pakistani youth are able to implement bilingual or even multilingual identities, thus meeting the needs of the local and the global communities. The mixture of the English, Urdu, and regional languages thus represents belonging to a global digital zone and also maintaining ties to the local cultural identities.

2.3 Digital Vernaculars and Sociolinguistic Identity

The interaction between language and identity in Pakistan is a very complex phenomenon, which could be explained by the colonial experience of the nation and the use of language as a tool of structuring social orders. English language is commonly identified with prestige, modernity, and high-quality education, and Urdu is considered to be a national identity, and local languages are a representation of ethnic and cultural affiliation. In its turn, this means that the linguistic choices of Pakistani young people on social media sites like Instagram could be used as the predictors of not only individual identity but the wider socio-political belonging.

The use of hybrid digital lexicons on Instagram e.g., one can mention the phrases like unfollow hogaya or attitude on hai is a demonstration of the need to negotiate multilingual and multicultural identities in the globalised context. These words of the lexicon are functioning as a tool of sociolinguistic negotiation thus showing how the youth is situated in the discourse societies: international and domestic. In his example Pennycook (2007) argues that language is not a rigid but a fluid system, a fact that is especially relevant in digital practices in which linguistic forms are constantly being transformed.

In Pakistan, the Internet, specifically the digital platform, like Instagram, exists as a space of performativity of identity negotiation. The artistic and, sometimes, playful examples of code mixing which can be found on these sites point to a leaky linguistic status which is inherent to youth culture. It is not only through the language behaviours of the newer generations who still use these media to convey feelings, opinions and types of personality that they determine their own identities but also conform to the standards of transnational and indigenous cultures.

2.4 Language and Technology: New Media Literacy.

The emergence of new literacies which come to be known as the new media literacy has been orchestrated by new social media platforms such as Instagram. This construct is the constellation of competencies that are needed to negotiate and communicate in the digital milieu, among the other capabilities, there is the ability to understand and create multimodal discourses, i.e. the integration of linguistic and visual modalities (Leu et al., 2015). In the Pakistani context, the role of new media literacy is growing because it determines the way the young population is interacting with the digital media and therefore is creating a new digital space like the Instagram that is leading to linguistic hybridity.

Social media serves as a platform of linguistic experimentation and innovation in the example of adolescents in Pakistan, where Instagram is used. The use of hashtags, visual effect, and emoticons are often used together with verbal expression in order to enhance the discourse and complement the interaction among the users. The lexical intermixing thus embodied in the hybrid digital lexicon that arises on Instagram is therefore not only inherently reliant upon lexical intermixing but is also inherently dependent on the multimodal affordances of digital platforms. The combination of text and visual elements and the supporting digital symbols generates a new form of communication that is contextual and highly innovative.

3. Methodology

The study is a corpus-based, mixed-method research paper, which will examine the trends in code-mixing and lexical innovations in Instagram posts by Pakistani users. It is analytical methodology relying on both quantitative linguistic methodologies and qualitative discourse analysis that enables to explore in a comprehensive way how language practices in digitally mediated spaces, in particular, on Instagram, can serve to express identity, emotion, and a feeling of engagement into local and global cultural environment. The research is based on the idea that young people are one of the most active population groups on social media platforms and uses Instagram as the social media platform, which is very popular in Pakistan.

3.1 Theoretical Framework

This paper is based on two broad theoretical approaches explaining linguistic behavior which could be observed in the corpus of Instagram: Matrix Language Frame (MLF) Model (Myers-Scotton, 1993) and performative identity concept in digital sociolinguistics (Bucholtz and Hall, 2005; Pennycook, 2007).

3.1.1 Frame Model of Matrix Language.

The MLF Model is one of the most important paradigms of explaining the language interaction on the level of clause, and it has been promulgated by Myers Scotton (1993). The model claims that a syntactic frame of a clause is provided by one language (the matrix language), and the other languages (the embedded languages) provide lexical items, e.g. substantive morphemes, e.g. nouns, adjectives and verbs. The theoretical construct might be especially useful in the interpretation of grammatical spellings of code-mixed utterances in the Pakistani Instagram corpus, where the Roman Urdu is habitually the matrix language, that is, the language that

offers the syntactic structuring, English is the language that offers the lexical input or clause of expression.

For example:

“Aj mood totally off hai.”

Kal hogaya kal assignment.

In these situations, Urdu is syntactically constructed as compared to English that offers content delivering an expressive or technical meaning. This is consistent with the finding of Awan et al. (2025) who confirm that a patriarchal functional relationship between languages dictates code-mixing in Pakistan on a regular basis.

3.1.2 Identity Performance and Digital Sociolinguistics.

As much as the MLF model is an excellent approach of examining the structural aspects that encircle the code-mixing, the model does not address the social roles of the language mixing. To that end, the given paper adheres to the model of performative identity construction, which was coined by Bucholtz and Hall (2005), and metrolingualism developed by Pennycook (2007). These models highlight the performativity and fluidity of the identity within the digital spaces. Bucholtz and Hall (2005) think that there is no such thing as identity but it is created and enacted in the continuous manner due to discourse and social interaction. Language is one of the arenas on Instagram where youths are acting, asserting, even negotiating identities often through a mixture of regional, nation-national, and global linguistic resources.

E.g. Instagram posts such as:

VIP package: No light + Karachi heat 🤔🔥.

#MoodOffHai after losing at PSL.

These posts are beyond the sharing of information, but it is identity performance, and the language is an expression of humor, frustration and place pride. The concept of metrolingualism proposed by Pennycook (2007) of blending and creative fusion of languages in urban life can be applied in the present case because it implies how Pakistani young people manage to deal with their multi-dimensional identity using the hybrid means of communication. In so doing, Roman Urdu and English are not modes of communication but of identity, identifiers of group membership and membership in a culture.

3.2 Research Design

The given work is based on a corpus-based paradigm, which is a mixed-method approach that combines quantitative linguistic analysis and qualitative discourse analysis to investigate the existence of code-mixing, lexical innovation, and the formation of identity in Instagram posts written in Pakistan. It is both an exploratory and interpretive research since it looks into the ways in which Instagram digital language practices can be used to learn more about the bigger sociolinguistic processes among Pakistani youth. One can also say that the affordances of

Instagram and, in particular, its visual and textual synthesis make it an attractive location to carry out such an analytical research.

The quantitative approach with the frequency and patterns of linguistic innovations and the code-mixed lexical items is evaluated, and the interpretation of the affective, identity-based and socio-cultural roles of such practices is done through a qualitative discourse analysis. The role of language in the roles of communication and the role of performance in an online environment can be easily observed due to the combination of these approaches.

3.3 Corpus Construction

It was created on a personalized corpus comprising of 1,200 publicly available Instagram posts collected in January 2024-June 2025. The sampling of the posts was done using their user accounts, which had the following qualifications:

- The self-identified location of the users was Pakistan.
- There existed English, Roman Urdu, Urdu script and mixed languages.
- Users used hashtags (e.g. #KarachiVibes, #PindiSwag), or geotagged places in Pakistan.

The sample was spread out using stratified sample by gender, urban-rural origin, and regions (Punjab, Sindh, Khyber Pakhtunkhwa and Baluchistan) in order to come up with a diverse sample in terms of language and cultural perspective. The mean user account posted approximately 30 times and this gave an equal measure of how the language was used by the different social groups. This stratification comes handy in reflecting the heterogeneity of the language practices in different parts of Pakistan in which norms and practices of the languages could vary.

Inclusion Criteria:

- Posts written in English, Roman Urdu, or bilingual/multilingual forms.
- Posts featuring hashtags, captions, comments, or story screenshots.

Exclusion Criteria:

- Posts from commercial brands or influencers with more than 100k followers.
- Posts that were meme-based, auto-translated, or unrelated to Pakistan's context.

3.4 Data Collection Tools

Several tools were used to collect and analyze the Instagram data:

1. Python-based Instagram Scraper: A custom scraper was developed using Selenium and BeautifulSoup to extract metadata (user ID, date, caption, hashtags, location) and text content from publicly available Instagram profiles.

2. Manual Annotation: Each post was manually annotated for language type, script usage, code-mixing patterns, and lexical features.

3. AntConc 4.2.0: This corpus analysis tool was employed for frequency analysis, keyword extraction, and collocate analysis, focusing on lexical innovation and bilingual patterns.

4. NVivo 14: This qualitative data analysis software was used to facilitate thematic coding and discourse analysis, particularly focusing on identity expression, emotional tone, humor, and other social functions of language use.

3.5 Ethical Considerations

This study adheres to the ethical guidelines set forth by the British Association for Applied Linguistics (2022) and the Association of Internet Researchers (AoIR). All data were collected from publicly accessible Instagram profiles, ensuring compliance with ethical standards:

Anonymization: Personal identifiers such as usernames, profile pictures, and comments tagging individuals were anonymized.

Security: Data were stored securely and encrypted during analysis.

Non-Interaction: No direct interaction with users occurred during data collection.

The corpus was constructed for academic purposes only and will not be published in full to ensure user privacy is upheld.

3.6 Analytical Framework

A dual-level analysis was conducted, incorporating both quantitative and qualitative methods:

3.6.1 Quantitative Analysis

Word Frequency & Keyword Analysis: This analysis aimed to identify common lexical innovations, hybrid terms, and borrowed structures. High-frequency items were examined in context to determine their function and script usage.

Code-Mixing Pattern Recognition: Posts were categorized according to Muysken's (2000) typology: Insertion, Alternation, and Congruent Lexicalization.

Hashtag Analysis: Code-mixed and stylized hashtags were cataloged for their emotional, cultural, or humorous functions.

3.6.2 Qualitative Analysis

Discourse Analysis: Selected posts were closely analyzed to explore the role of language in identity construction, emotional expression, and group affiliation.

Thematic Coding: NVivo-assisted coding identified dominant themes such as humor, emotional distress, national or regional pride, and fashion/trend signaling. These themes were then analyzed for recurring lexical and stylistic patterns, focusing on the affective and social functions of language mixing.

3.7 Code-Mixing Typology

The typology used for coding code-mixing in the corpus is summarized in the table below:

Type	Definition	Example (from corpus)
Insertion	A lexical item from L2 embedded into L1 syntax	“Aaj ka plan cancel hogaya, total disaster!”
Alternation	Clause-level switching between languages	“We went to Murree. Thand amazing thi wahan”
Congruent Lexicalization	Lexical items from both languages within a shared syntactic frame	“Exam stress on hai but still chilling with squad”

Code-Mixing Typology

The following table presents the typology used to classify code-mixing patterns in the corpus, tailored to reflect the linguistic dynamics commonly observed among Pakistani Instagram users. The typology helps to identify the structural and functional aspects of code-mixed expressions, capturing how various languages—primarily Urdu, English, and Roman Urdu—are blended to create social meaning in the digital space.

Type	Definition	Example (from corpus)
Insertion	A lexical item from one language (L2) is embedded into the syntactic structure of another language (L1). Typically, L1 maintains the grammatical framework, while L2 provides the content. “Aaj ka plan cancel hogaya, total disaster!” (Urdu syntax with English content, common in daily posts when discussing events or plans)	
Alternation	Switching between two languages at the clause level, where each language operates within its own syntactic unit. This often occurs when shifting topics or expressing complex ideas. “We went to Murree. Thand amazing thi wahan.” (A blend of English for introducing the idea and Urdu to describe the experience)	
Congruent Lexicalization	Lexical items from both languages are inserted within a shared syntactic frame, where both languages contribute to forming the meaning without strict syntactic boundaries. “Exam stress on hai but still chilling with squad.” (Combining English technical terms with colloquial Urdu expressions to create a relaxed yet academic tone)	

Contextual Relevance to Pakistan

The most observed pattern in the Pakistani context is the code-mixing of the English and Urdu languages whereby Roman Urdu the Urdu written with the use of the Latin alphabet acts as the linguistic link between the two languages. The hybrid register is especially common among the younger social media users of platforms like Instagram, where informal, creative, and localized forms of expression are important tools of communication.

The most common form of insertion is in urban areas where there is regular use of English in education, business and pop culture. Service terms like total disaster or just chillin are commonly incorporated in otherwise Urdu sentences, which in turn signifies the domination of English influence in online communication especially in metropolitan cities like Lahore, Karachi and Islamabad.

An example of the bi-lingualism character of most Pakistanis is alternation, which is where in a discourse or a post, there is a fluent alternation between Urdu and English. It seems to be especially acute in the posts made by young people who write about activities, feelings, or social events, where the combination of English and Urdu creates a more expressive and culturally sensitive conversation.

The congruent lexicalization phenomenon is a kind of a vehicle to express the modern or trendy ideas, particularly in the digital milieu. The lexical elements like squad, stress, chilling is in coexistence with the Urdu lexical elements, hence a phenomenon of hybridization of native and foreign linguistic resources. These multimodal statements resonate with Pakistani teenagers who strive to reconcile between belonging to their native culture and being part of the global cultural inputs.

The typology was analyzed along with the variables like script use (Roman Urdu, Urdu script, English), spatial difference (e.g. differences in code-mixing between Punjab, Sindh, Khyber Pakhtunkhwa, and Balochistan), gender (to study possible differences in linguistic expression between men and women), and affective register (emotional valence, humor, sentiment). Comparative studies of these dimensions showed less obvious patterns of discourse construction and, thus, provided information about how Pakistani youths negotiate identity and engage in digital culture through language.

4. Data Analysis

In this section, the data-analysis steps that are used to examine the language habits of Pakistani Instagram users are outlined, especially focusing on the areas of code-mixing, lexical innovation, and identity formation. The analytical framework incorporates the quantitative and the qualitative paradigm, hence the consideration of complexity of language use in the multilingual digital environment in Pakistan. The general aim is to explain trends of code-mixing, interaction between languages and the emotive and social functions of the linguistic strategies in the background of Pakistani youth culture.

5.1 Quantitative Analysis

The quantitative part of the analysis was guided by the determination of salient trends in the language used in posts on Instagram. The main analytical tools were the frequency analysis tool Antconc version 4.2.0 and Python-written web scrapers that could be used to extract the

relevant metadata and textual data. In this analysis, various dimensions were involved; the code-mixing, lexical innovation, and the use of Roman Urdu, Urdu script, and English in the corpus.

5.1.1 Word Frequency & Keyword Analysis

The first stage of the quantitative analysis involved determining the frequency of code-mixed words, neologism, and frequently borrowed words of expression used by Pakistani youth in Instagram. This stage of the analysis suggested a marked predominance of English loan words in conjunction with Urdu, with words such as ‘mood off’, ‘chilling’, ‘attitude’, and ‘unfollow’ being original examples of the words commonly used. These words were often used in the sentence structure of Roman Urdu, showing how the youth cleverly blended local with international languages.

The highest-frequency items identified in the analysis were:

Mood off (mood off hai)

Chilling (chilling with squad)

Attitude on hai

Pindi swag

Unfollow hogaya

These terms reflect the hybrid digital lexicon that Pakistani youth engage with on Instagram. Words from English, often related to emotional expression, trends, or technology, are integrated into Urdu sentences to create a more dynamic, expressive form of communication.

5.1.2 Code-Mixing Pattern Recognition

Using Muysken’s (2000) typology, the posts were tagged based on the pattern of code-mixing. The three types of code-mixing identified in the corpus were:

Insertion: English terms inserted into the syntax of Roman Urdu or Urdu representing emotional and technical values.

E.g. “Aaj ka plan cancel hogaya, total disaster!” (English words such as “total disaster” have been inserted into the Urdu structure in order to represent the negative tone of the post.)

Alternation: Switching languages at the level of clauses. This switching took place in posts particularly when there was a switching of topic or tone.

E.g. “We went to Murree. Thand amazing thi wahan” (The clause “We went to Murree” is in English, while “Thand amazing thi wahan” is in Urdu as it describes the local experience.)

Congruent Lexicalization: Lexical items from both languages are inserted into the same syntactic frame.

E.g. “Exam stress on hai but still chilling with squad” (The mix of English and Urdu form an idiomatic expression of the stress felt translating into the relaxed social engagement which follows)

In order to realize how these forms of code-mixing come about in the everyday discourse of social media, the frequency of each of the forms of code-mixing was calculated. In this corpus, insertion was the largest number of types of code-mixing, followed by alternation and congruent lexicalization.

5.1.3 Hashtag Analysis

The analysis of hashtags was integral to understanding the cultural and social functions of code-mixing. Hashtags in the corpus reflected the same linguistic blending observed in captions and comments. Commonly used hashtags included:

#KarachiVibes

#PindiSwag

#MoodOffHai

#SwagLife

#PSLSeason

These hashtags often combined English and Urdu to convey local identity, pride, and humor. For instance, #KarachiVibes reflects the global trend of hashtagging while embedding local cultural identity in a way that is recognizable both within and outside Pakistan.

5.2 Qualitative Analysis

The qualitative analysis carried out through NVivo 14 which was used for thematic coding and discourse analysis focused on the understanding of the social, emotional, and identity feeling behind code-mixing. The close reading of the selected posts led to a identification of patterns which indicated how Pakistani youth use language on Instagram in order to express identity, affiliation and emotion.

5.2.1 Identity Construction

A key finding from the qualitative analysis was the fact that language on Instagram is used as a performative resource for identity construction. The framework of performative identity (Bucholtz & Hall, 2005) was also used for the analyses of how language is also used to convey and express social identity. Instagram was seen to be a site in which various aspects of identity are negotiated i.e. regional, national, and global. Regions reflected in hashtags like #PindiSwag or #KarachiVibes exemplify a very strong regional identity affiliation with the regions of Rawalpindi and Karachi. These expressions represent a pride in the local culture and an alignment with larger social and cultural movements. The making use of hybrid expressions by users creates a performance of local identity also with a global awareness. The expressions of modernity and youth culture embodied in English are mixed with expressions pointing to local pride.

5.2.2 Emotional Expression and Humor

An additional important function of code-mixing in the results was its use towards emotional expression. The code-mixed phrases “Mood off hai” and “Unfollow hogaya” indicate the sentiments of frustration, humor or social critique. The English terms “mood” and “unfollow” give the message a sarcastic or humorous edge which is emphasized by the local context of the feelings expressed with the Urdu words. These hybrid expressions allow for a range of emotional choices, which appeal to a youthful audience.

Some sample posts include:

“#MoodOffHai after watching PSL loss.”

“VIP combo: No light + Karachi heat 😎🔥”

do reflect not only the emotional distress (frustration caused by a sports loss or weather-related complaints) but also sense of humor which is often to be found in online expressions of youth in Pakistan. The use of Roman Urdu itself provides for intimacy, since it carries a conversational tone which resonates with peers, while it maintains a connection to local identity.

5.2.3 Social and Cultural Function of Language Mixing

Social function of the linguistic strategies used in Instagram also seems to be of prime importance, where language mixing serves to perform group membership and signal affiliation. The data here also show that the hybrid language forms that are employed allow users to affiliate with varying social circles and groupings depending on the regional identity, youth culture or shared experiences. The use of code-mixing signals a conscious participation in worldwide youth culture while still maintaining links to local customs and cultural practices.

For instance, posts about different sports like references to the Pakistan Super League (PSL) or aspects of national pride, etc., were sometimes expressed with a mixture of English and Urdu so that followers might experience traditions of the people in important local ways. English was preferred in posts for global appeal and Urdu was used for emphasis at the local level. In this way youths are able to resolve some of the difficulties of localism versus globalization.

5.3 Regional and Gender Differences

The last layer of analysis investigated differences in code-mixing patterns across regions and gender. While the similarities in hybrid language forms was considerably evident, there were slight divergences detected. For instance, the urban users in Karachi and Lahore often employed a more frequent pattern of English use and a greater dominance for globalized slang than rural users. Differences between the genders were evident too in the way humour, irony and tone emotional were expressed: the male users displayed a more relaxed, confident tone, whilst the female users tend to express their sense of humour with indications of affection or self-deprecatory humour.

Discussion

The analysis presented here shows that code-mixing with English and Urdu on Pakistani Instagram is a highly social and organized linguistic process. It has been shown in this study applying Bucholtz and Hall’s (2005) performative identity framework that Instagram users are performing identity work by making choices about the language they use. The fact that Roman Urdu is used in a flexible and innovative way illustrates the potential of the vernacular in Pakistan's digital context. The non-standardized nature of the language in terms of institutions

and educational bodies means that these linguistic choices allow for a more accessible means of expression, especially by youth, than if they were to use the canonical varieties of Urdu and English which tend to be used in much more formal environments.

The data illustrates how Pakistani Instagram users are engaged in essentially multilingual language practices - exploiting the syncretism of English, Urdu and the regional languages of Pakistan in the performance of identities in their online persona. Roman Urdu born especially in the digital context acts as the bridge between these languages along which the performance of identities can be undertaken using aspects of both global and local identities. As Blommaert and Rampton (2011) have shown, digital language use is shaped by a mix of local ideologies and global flows. In the context of Pakistan this means that English is part of a colonial legacy and also a marker of modernity and social mobility, while Urdu has as a national language an emotive and cultural resonance.

The combination of humorous references, neologisms, and local slang shows the participation in constructing the norms of informal communication by the digital users in a participatory culture. The manipulation of language use by the youth of Pakistan also contributes to the global discussion about digital multilingualism. They are active participants in putting forward new norms of language use, instead of just being passive audiences of global digital culture. This work (Pennycook 2007) suggests that certain globalized digital practices, such as code-mixing and hybrid language forms are relevant to the construction of identity in online locations.

This research may contribute to the prevailing discussions about multilingualism as it relates to the youth semiotics, especially in terms of how language use in the digital spaces goes beyond mere communication. Language serves as a vehicle for social meanings as well as identity performances and cultural participation. The function of code-mixing and neologisms in digital areas, especially in social network sites such as Instagram, also shows the fluidity of language in Pakistan's socio-linguistic context which is changing.

The analysis of the data has shown that code-mixing and lexical innovation on the Instagram platform in Pakistan are not only linguistic phenomena but also important means to perform identity and express their emotions. The hybridization of the English and Urdu language presents a situation in which young people are negotiating their multilingual identities, and the youth are caught in the middle of the struggle between their local cultures and the global digital. In the analysis of the hashtags, expressions of emotions and identity constructions, the social and cultural role of these language practices in the digital vernacular of Pakistan can be explained where language is applied creatively to create social meaning and identify personal and group identities.

This paper has demonstrated how Pakistani Instagram users operate within dynamic and context specific language practices that adhere to code-mixing and lexical innovation of combining English, Urdu, and regional terms. Instead of indicating a linguistic deficit, the practices indicate creative mastery of various linguistic repertoires. In such hybrid forms, Pakistani young people form identities, indicate emotion, and engage in digital culture.

Dominance of Roman Urdu and stylistic features of the hashtags and the introduction of the bilingual neologisms all evidence a shift in the language environment in Pakistan, which disrupts traditional views of standard language. These results imply some significant implications on language education, sociolinguistics and the policymaking:

Language Education: Inclusion attitudes towards digital literacy need to be introduced in the education systems of Pakistan. The research shows that young people are proficient in digital literacy, inventing novel modes of communication that goes beyond the school educational standards.

Sociolinguistics: This research calls for a redefinition of vernaculars in the context of social media. The blending of multiple languages on platforms like Instagram challenges traditional sociolinguistic categories like L1 vs. L2 and standard vs. non-standard language.

Policy-making: Digital communication in Pakistan needs to be recognized as a legitimate site of cultural and linguistic production, with attention to the role of informal digital communication in shaping communicative norms.

6. Implications and Future Directions

The findings of this research have significant implications for education, sociolinguistics, and digital policy-making. They also open several avenues for future research to explore:

6.1 Educational and Digital Literacy

In Pakistan, traditional education systems tend to marginalize non-standard forms of communication like Roman Urdu and code-mixed English, often labeling them as “incorrect” or “slang.” However, the study suggests that these forms of communication reflect high levels of linguistic creativity and digital fluency. As Pakistan continues to integrate more youth into the digital world, curricula should evolve to include these practices. Schools and universities could benefit from encouraging students to engage critically with digital language practices, bridging the gap between classroom norms and the everyday digital communicative realities of youth.

6.2 Sociolinguistics and Language Planning

From a sociolinguistic perspective, this research challenges the dichotomy of formal vs. informal language use. Roman Urdu and code-mixing are not indicative of linguistic decline but instead represent active, meaningful, and creative language use. Language planners and policymakers in Pakistan must take these forms of communication into account when developing language policies. Incorporating digital language practices in policy-making will help recognize the role of informal communication in shaping linguistic and cultural norms.

6.3 Research and Policy

The study has some policy implications, especially concerning the language rights and digital inclusion. Informal digital communication should be valued by public institutions and particularly when it comes to youth interaction. The study suggests the need to investigate more the role of regional languages (e.g., Pashto, Sindhi) in digital code-mixing and the way they are combined with the global languages such as English and Urdu. Moreover, the attention of

research should be paid to differences in digital literacy based on classes and how the trends of digital behavior can affect online, as well as off-line, discourse.

Further research also needs to move beyond Instagram and consider platforms such as Tik Tok, WhatsApp, and other platforms that use social media to have a wider perspective on how the Pakistani youth culture applies the language. Moreover, the study of the neologism development and its influence on the offline communication would provide useful information on the processes of the language change in the multilingual context in Pakistan.

7. Conclusion

To sum up, this paper gives an in-depth examination of the manner in which online language practices are practiced by Pakistani Instagram users in a creative manner. It points to the imaginative and mobile nature of how young people negotiate their identities by mixing codes and using lexical innovation and performative speech acts. Instagram in Pakistan is, in fact, more than a mere tool of communication, and a linguistic playground, where young people in Pakistan are actively reimagining and redefining multilingualism in their own ways. The implications of the findings on language education, sociolinguistic theory, and the digital policy are more comprehensive and would signify that informal digital communication must be considered an important part of the developing linguistic situation in Pakistan.

References

- Awan, F., Mehmood, S., & Khan, S. (2025). *Bilingualism and identity in urban youth: A case study of Lahore*. Lahore University Press.
- Blommaert, J., & Varis, P. (2015). *Language and the digital age: Sociolinguistics and digital communication*. Routledge.
- Butt, M., Iqbal, M., & Abbas, M. (2025). *Language practices of bilingual youth in Pakistan: Insights from social media*. Cambridge University Press.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.
- Jabeen, S., & Memon, S. (2019). Multilingualism in Pakistan: A sociolinguistic analysis. *Journal of South Asian Linguistics*, 2(1), 45-59.
- Leu, D. J., Kinzer, C. K., Coiro, J., & Cammack, D. (2015). New literacies: A dual-level approach to the study of language and communication. *Educational Psychology Review*, 27(4), 583-599.
- Mansoor, S. (2005). *Language and education in Pakistan: The role of English in bilingual classrooms*. British Council.
- Pakistan Telecommunication Authority. (2025). *Internet penetration and digital trends in Pakistan*. Government of Pakistan.
- Pennycook, A. (2007). *Global Englishes and transcultural flows*. Routledge.