

Banglалish on social media: A corpus-based study of Bangla-English code-mixing across four platforms in Bangladesh

SHAJADUL ALAM SWEET¹

MD. NURUL KABIR EMON¹

SADIA KHANDOKAR MIM¹

FAKHRUL ISLAM MAHIM¹

MAHFUJ HOSEN¹

IMRANUL HAQUE MASUM²

¹The University of Asia Pacific

²The Islamic University

Abstract

This article reports a 12-month, 120,000-token corpus study of Bangla–English code-mixing in Bangladesh’s social-media discourse across Facebook, Twitter, YouTube, and Instagram. Anchored in variational sociolinguistics and contact linguistics, we operationalise code-mixing using the Matrix Language Frame (MLF) model and related structural diagnostics (insertion vs. alternation; borrowing vs. switching), and we quantify platform effects on mixing frequency and structure. Findings indicate stable insertional patterns with Bangla as the matrix language; productive hybrid verb constructions (e.g., *comment korchі, share korbo*); strong preferences for Bangla determiners with English content nouns (e.g., *ei app, amar phone*); and domain-specific borrowing of digital-culture vocabulary. Platform affordances condition mixing: Facebook supports longer mixed turns, while Twitter compresses mixing and favours hashtag innovations; YouTube concentrates emphasis and evaluation in comment bursts; Instagram integrates visual–textual cues with searchable, multilingual tags. Demographic correlates (age, education, gender) are treated as tendencies rather than fixed categories. The article contributes (i) a transparent, replicable coding scheme for code-mixing in South Asia, (ii) cross-platform evidence of code-mixing as structured, not ad hoc, and (iii) research/pedagogical/policy implications of treating code-mixing as a regular, rule-bound resource in Bangladesh’s digital communication ecology.

Keywords: code-mixing; code-switching; Bangla–English; social media; Matrix Language Frame; borrowing; insertion; alternation; corpus linguistics; Bangladesh

Introduction

The Digital Revolution of Bangladeshi Multilingualism

Every day, millions of Bangladeshi social media users move between Bangla and English in ways that no one could have imagined just ten years ago. When a university student posts “*Exam er pressure e ami completely broken hoye gechi*” (I have become completely broken under exam pressure), they are not simply switching between languages. Instead, they are helping to create a brand new type of language. This new language shows the complex realities of what it means to be Bangladeshi today.

This language mixing, which we call “Banglailish,” means much more than just casual language blending. It illustrates what Androutsopoulos (2015, p. 188) refers to as the “networked multilingualism” of digital communication. In this era, digital informal language learning has become an important trend (Liu et al., 2024). Technological advancements and global connections are transforming traditional language boundaries in society. Bangladesh has over 120 million internet users and a rapidly growing digital system. Social media platforms have become testing grounds for postcolonial multilingual expression. People use multiple languages in combination, reflecting their colonial history and current global connections.

Theoretical Positioning: Beyond Traditional Frameworks

Traditional World Englishes theory provides a good foundation. However, it struggles to explain how digital language varieties are formed today (Sergeant & Tagg, 2014). Kachru’s (1985) Three Circles Model placed Bangladesh clearly in the Outer Circle. In this circle, language varieties develop their own rules through contact with local languages. However, today’s digital practices show us a much more complex reality. Users now actively create and work out language rules through online interactions. They do not simply inherit these rules from schools and official places.

Recent research in digital sociolinguistics shows we need new theoretical frameworks. These frameworks must explain how platform affordances shape systematic linguistic change while creating entirely new possibilities for linguistic innovation (Thurlow & Mroczek, 2011). Platform features, such as character limits, emoji options, and messaging formats, influence how language changes over time. They also open up entirely new ways for people to be creative with language.

The present study adds to this growing field of research. It demonstrates how Bangladeshi digital communities collaborate to develop intricate, multilingual practices. These language practices serve several important purposes. They help people form their identity online, build solidarity within communities, and also help people negotiate cultural meanings and relationships in digital spaces.

The Postcolonial Digital Context

Bangladesh has been a crossroads of different languages for centuries, resulting in a language situation that is a blend of many languages. Muslim rule introduced the influences of both the Persians and the Arabs. It was British colonisation that made English the dominant language. The current digital transformation is a new phase in this contact history. Instead of passively adopting English as it is presented to them by others, young Bangladeshis have now begun to redefine English by redefining it themselves (Canagarajah, 2012). This is an essential democratisation process in the light of the recent political change in Bangladesh. In August 2024, a mass student-led uprising forced Prime Minister Sheikh Hasina to resign on August 5. Nobel laureate Muhammad Yunus was sworn in as chief adviser of an interim government on August 8 (Al Jazeera, 2024).

In the political and social context, language innovations in Bangladesh can be viewed as more than just language innovations. It presents the practice of the people of Bangladesh and how the youth are taking ownership of power and control in crafting their multilingual future. They do not want to wait until others tell them what they can do with language. They are instead seeking alternative means of expression. The presence of such an invention of language demonstrates that they are people who are independent and self-governed. That is to say, it means that they cannot be compelled to follow the rules of language; they are not subjugated by external power or even by a person who is superior to them.

Research, Innovation, and Scope

The paper fills an important gap in the literature on the role of digital technologies in the content provision of multilingual settings and ideational shifts in the South Asian linguistic landscape (Lee, 2017). Nevertheless, researchers have yet to systematically analyse these new forms of language that emerge in the online environment. This field is understudied and warrants closer examination. Our research has three innovative findings that dispute known assumptions about the mixing of languages.

First, Banglalish is far better grammatically organised than traditional models of code switching expected it to be. Language does not consist of an arbitrary combination, but it has definite patterns and rules (García & Li, 2014). Second, platform-related linguistic accommodations are demonstrating an unprecedented degree of complexity in multilingual digital competence. Facebook, Twitter, Instagram, and other platforms are highly skilled in ways that users adapt their language. Third, our results contradict deficit models according to which language mixing is a manifestation of imperfect language acquisition (Blommaert, 2010). These deficit models falsely imply that individuals confuse languages since they are not yet completely mastered in both languages. Instead, Banglalish plays high-level identity and community roles that facilitate the local cultural preservation and the global inclusion. Users strategically use this mixed language variety to articulate complicated cultural identities, develop community relations, and engage in local and global discourses at the same time.

Research Questions

This study addresses four interconnected questions:

- What systematic linguistic innovations characterise Banglalish across different digital platforms, and how do these patterns challenge traditional models of contact-induced change?
- How do platform affordances and user demographics interact to shape distinct varieties of digital multilingual expression?
- What identity and community functions does strategic language mixing serve for different user groups, and how do these functions vary across digital contexts?
- What implications do these findings hold for reimagining multilingual education and language policy in postcolonial digital societies?

Literature Review

Classical Theories of Code-Switching and Code-Mixing

Research on bilingual discourse has traditionally focused on structural models that attempt to explain the grammatical constraints governing code-switching and code-mixing. Poplack's (1980) seminal work introduced the Free Morpheme Constraint and the Equivalence Constraint, proposing that code-switching tends to occur at points where the surface structures of both

languages align syntactically, and not within bound morphemes. This early constraint-based perspective framed code-switching as a rule-governed practice rather than evidence of incomplete bilingual competence.

Muysken (2000) created a code-mixing typology that identified two types: insertion (the use of the syntactic frame of a language to embed lexical items of a different language), alternation (the use of two languages to alternate between structures), and congruent lexicalisation (the use of grammatical structures across two languages). This model stressed typological distance in languages to form mixing patterns and has been extensively used on bilingual populations.

A more comprehensive, empirically based framework was provided by the Matrix Language Frame (MLF) by Myers-Scotton (1993, 2002). The MLF model suggests that in mixed clauses, the linguistic content is supplied by the language that has the grammatical framework and system morphemes (the matrix one) and content morphemes (the embedded one). This model was further worked into the 4-M model (Myers-Scotton & Jake, 2017), which classifies morphemes based on the idea they activate and grammatical function. The asymmetric tenets of the MLF model have been validated in extensive corpus-based research in Spanish-English (Callahan, 2002), Turkish-German (Backus, 1996; Çetinoğlu, 2016), and Welsh-English (Deuchar, 2006) bilingual language communities.

Although these classical models still form the basis of grammatical regularities identification, they were constructed to explain the spoken discourse face-to-face. They fail to propose comprehensively the multimodal, platform-mediated, and identity-driven processes that typify language use in digital spaces. This restrictiveness has led to the creation of more recent theoretical frameworks that are more appropriate in terms of analysing online multilingual practices.

Emerging Theories of Digital Multilingualism

Recent sociolinguistic writing has left behind the idea of multilingual practices as a switch between discrete codes. It posits it as a fluid, multimodal repertoire that is influenced by digital technologies. Among the major frameworks is translanguaging, which redefines the concept of bilingualism in using unified systems of linguistic and semiotic resources, as opposed to using two distinct languages alternately (García & Li, 2014; Li, 2018). Li (2018) emphasises the idea that translanguaging enables speakers to create meaning, construct identity, and negotiate relations through the creative mobilisation of all available resources. This point of view is closely related to the Bangladeshi phenomenon, according to which the young generation of Bangladeshi people mixes the Bengali language, English, emojis, memes, and hashtags into the same expressive system.

Androutsopoulos (2015) proposes the concept of networked multilingualism, which refers to how the affordances and interactional practices of digital platforms shape multilingual practices. Lexander and Androutsopoulos (2023) develop this by theorising medial repertoires; the cross-modal semiotic resources of the users (text, image, audio, hashtags) which are coordinated according to the platform constraints, e.g., character limits or algorithmic visibility. This point of view is especially applicable to the case of Banglalish because the current study indicates that, in response to the affordances of Facebook, Twitter, YouTube, and Instagram, users change their mixing strategies.

The notion of superdiversity has influenced recent digital sociolinguistics (Vertovec, 2007), which emphasises the increase of linguistic, cultural, and semiotic diversity in online communication. According to Varis and Wang (2015), digital spaces can be defined as spaces of superdiverse repertoires in which users integrate kinds of resources to run identity and affiliation. Deumert (2014) also argues that mobile communication promotes individualised, multimodal repertoires that have disrupted the traditional language boundaries. These views contribute

to elaborating the extreme richness of the visual, textual, and interactional characteristics of Banglalish.

Recent research also provides an example of the influence of digital affordance and multi-modal resources on multilingual practices. As an illustration, Ma and Li (2024) illustrate how Chinese social media users use translanguaging strategies of mixing linguistic, visual, and interactional resources to produce online refusals. In contrast, Tao et al. (2024) reveal that when interacting online, international students combine multimodal repertoires—text, visuals, and audio—to convey the message. These results indicate the functioning of translanguaging on more than a cross-linguistic but on a cross-mode platform, which supports the importance of a multimodal approach in the analysis of Banglalish.

Building on this line of inquiry, recent studies in Digital Applied Linguistics (DAL) demonstrate how digital technologies and AI-based tools can enhance learner motivation and self-directed engagement in multilingual contexts. Huang and Mizumoto (2024) showed that integrating generative AI into writing tasks increased students' motivation and confidence in their writing abilities. Likewise, Huang and Teng (2025) showed that integrating GenAI increased learners' engagement. Dizon et al. (2025) reported that university EFL learners used ChatGPT to support their self-regulated learning, enabling them to practice language more independently beyond the classroom. These findings reinforce the view that digital platforms can play a constructive role in shaping multilingual practices, aligning with our interpretation of Banglalish as a dynamic, agent-driven linguistic repertoire.

The idea that languages exist as separate, fixed entities has been criticized, especially in post-colonial studies, which highlight how power shapes global Englishes. Makoni and Pennycook (2006) argue for the disinvention of languages as social objects by examining them in their constructed form. Pennycook (2017) focuses on the process of English appropriation and recontextualisation on the local level, whereas Canagarajah (2012) emphasises that multilingual speakers in postcolonial locations reuse English to demonstrate agency. Such structures offer a valuable perspective on the interpretation of Banglalish as a decolonial practice that enables the Bangladeshi youth to recover the ownership of English in the cyber world.

Integrating Structural and Poststructural Perspectives

Although the classical models can be used to explain the grammatical structure of code-mixed discourse, new frameworks can be used to explain how social and technological practices underlie the production of such a discourse. The MLF model is still helpful to determine the constant Bangla matrix forms and the prevalence of system morphemes in Banglalish. Nevertheless, the translation in language theory explains the deployment of these structures as components of more comprehensive identity acts. In contrast, networked multilingualism explains how platform-specific affordances influence the mixing practices.

Consequently, this paper assumes a synthesised approach: Banglalish can be seen not only as (i) a systematised linguistic system that follows the rules of an MLF model but also (ii) as a dynamism represented by translanguaging, networked multilingualism, and superdiversity. This twofold approach enables a more detailed understanding of how Bangladeshi social media users utilize the affordances of the platforms to develop hybrid identities, negotiate cultural meanings, and introduce new forms of multilingual expression.

The study helps to fill in a serious gap in the existing literature. Earlier research has been inclined to analyse the mixing of Bangla-English on structural scales such as the Matrix Language Frame (MLF) model or on sociocultural approaches such as translanguaging and networked multilingualism. Nonetheless, none of them have integrated these methods to describe how such mixing works in the digital environment of Bangladesh. This research is a novel

combined theoretical approach that explains both the grammatical roots and the identity-based creativity underlying Banglalish by uniting structural and poststructural approaches.

The latter aligns with the recent claims by Hamid et al. (2024) that applied linguistics research in Bangladesh needs to shift beyond descriptive research and contribute to building a global theory through the Global South. Placing Banglalish in the context of this agenda, the current study will make an additional contribution, not only by providing empirical evidence but also by offering a theoretical framework based on the sociolinguistic realities of the digital ecology in Bangladesh.

Switches can happen under the Classical Constraints model (Poplack, 1980; MLF), but this model does not explain why people use them differently on different platforms. Our digital corpus shows identical frames redistributed by affordances (see Table 3), which these models do not predict, motivating an affordance-aware expansion. Methodologically, our coding scheme and cross-platform corpus extend corpus-based social-media work (Rüdiger & Dayter, 2020) as a replicable template for South Asia.

Methodology

Research Design Philosophy

This investigation uses an innovative mixed-methods approach that combines computational corpus analysis with ethnographic sensitivity to user practices. Rather than imposing external categories on digital multilingual practices, our methodology prioritises what people use multilingual digital practices to accomplish in their everyday social interactions (Tagg, 2015). The research design showcases recent advances in sociolinguistic methodology, focusing on remote data collection techniques adapted to digital environments (Rymes & Leone, 2014). Our approach recognises that smartphone and computer technology have advanced to enable high-quality linguistic research while respecting users' privacy and consent practices. We can now collect good data without invading people's personal space or violating their trust.

Corpus Development Strategy

The Bangladeshi Digital Corpus includes 120,000 tokens systematically collected from four major social media platforms: Facebook (40,000 tokens), Twitter (30,000 tokens), YouTube comments (30,000 tokens), and Instagram (20,000 tokens). This distribution shows actual usage patterns among Bangladeshi social media users. Facebook serves as the primary platform for extended discourse. Instagram represents emerging visual-textual integration practices.

We located public posts via platform search interfaces using a bilingual seed list of 142 keywords/hashtags (Bangla script and Latin transliterations), expanded monthly by co-occurrence mining. We excluded organisational/brand/verified accounts, as well as reposts/retweets, and likely bot activity (≥ 5 posts/min sustained or duplicate content across ≥ 3 accounts). Per-user contributions were capped at 15 posts/month to avoid over-representation. Script detection and language ID filtered Bangla/English content and Latin-script Bangla.

Data collection spanned 12 months (January–December 2024) to capture temporal variations and seasonal linguistic changes. This extended timeframe was crucial given Bangladesh's political transitions in 2024. These transitions significantly influenced public discourse patterns and language choices across digital platforms. The extended collection period enabled researchers to observe how major political events influenced the way people used language online. These political events formed part of the discursive backdrop; while we do not claim causal effects, our 12-month window ensures coverage of contemporaneous shifts in public talk.

Participant Demographics and Sampling

The study focuses on urban Bangladeshi users aged 18–45 from Dhaka, Chittagong, and Sylhet metropolitan areas. Stratified sampling ensures representation across three age groups (18–25, 26–35, 36–45), three education levels (secondary, tertiary, postgraduate), and gender balance. Only public posts were included, in accordance with established ethical guidelines for social media research.

User identification relied on profile analysis, network connections, and linguistic markers while ensuring complete anonymisation of personal information. The geographic focus on urban contexts reflects the reality that English–Bangla mixing is most common in cities. These cities have high internet penetration and broad access to English-medium education. Rural areas often lack access to internet services and English education, which means that the use of such language mixing is not as common.

Linguistic Analysis Framework

The use of Banglalish language is explored in four aspects. Such lexical patterns as loanwords, loan translations, semantic extensions, and new compound structures are identified. The morphosyntactic patterns emerge in the hybrid forms of the verb phrases, that is, the combination of noun phrases and inflectional adjustment procedures. In particular, orthographic innovations are emphasised, as they reflect the specific challenges of encoding Bangla phonology on a digital surface using the Latin script. Functional roles (including pragmatic functions) are examined systematically in terms of identity construction, solidarity marking, humour patterns, and register variation.

Manual annotation is performed according to pre-defined coding schemes based on contact linguistic structures (Backus, 2003). Special emphasis is placed on channel-specific features, such as hashtags, mentions, and emojis. Several trials are conducted to assess the inter-rater reliability of the annotators, ensuring that all data are analysed at the same level by the same annotators. Three annotators double-coded 20% of the corpus (24,000 tokens). Cohen's $\kappa = 0.78$ (lexical types), 0.82 (hybrid verbs), 0.76 (functions); mean agreement 89%. Disagreements were adjudicated in weekly meetings. Such testing is designed to ensure that multiple researchers review the same data and use the same code, thereby increasing the credibility of the results.

Platform-Specific Analysis Protocols

Different platforms require a specific method of analysis, as their communication strategies differ. Facebook can be analysed as a medium for long monologues, status updates, and group discussions, allowing for sophisticated linguistic experimentation. While Twitter-based data mining focuses on novelties based on brevity, hashtag inventiveness, and trends in real-time reactions, the analysis of YouTube comments concentrates on the discourse and patterns of reaction words. Analysing Instagram involves examining visuals and text, as well as utilising visual aesthetic language.

A cross-platform comparison will allow us to highlight general trends and platform-specific adaptations. This tells us something about the role of technological affordances in systematic linguistic change (Boyd & Ellison, 2007). The comparison between the different platforms will help the researcher identify the common and distinct patterns of language that exist across all platforms. This illustrates the evolution of the language over time, as influenced by the platform's capabilities.

Quality Assurance and Validation

Specific validation methods were maintained to ensure the quality of the methodology. The inter-rater reliability was assessed by the three trained annotators independently annotating

20% of the corpus. Conflicts were resolved through discussion. Member checking involved presenting interim findings to a group of participants to ascertain the accuracy of interpretation. Quantitative pattern analysis was therefore triangulated with a qualitative analysis of discourses to present a comprehensive overview of usage functions.

These verification methods ensure the authenticity and accuracy of the study's results. The study results will be more credible if multiple researchers reach the same conclusions using the same data. The presence of real participants can help verify that researchers accurately understand how real users use words in a particular way.

Results: Revolutionary Patterns in Digital Multilingual Practice

Lexical Innovation: Beyond Simple Borrowing

The lexical analysis reveals word-choice processes that extend beyond simple borrowing hierarchies. While technology vocabulary predictably favours English insertions (e.g., *internet*, *app*, *smartphone*, *download*), unexpected patterns emerge in emotional and social domains. Young users tend to use English terms systematically, even in cases where perfectly matching Bangla terms exist: *excited hoye gechi* (I have become excited) rather than *uttejito hoye gechi*, or *nervous lagche* (I am feeling nervous) instead of *udvigno lagche*. This implies that word choice serves identity performance rather than communicative efficiency.

The preference for English in stance-taking ("*excited hoye gechi*", "*depressed mone hocche*") reflects the symbolic hierarchy between the two languages. English indexes modernity, professionalism, and emotional distance, while Bangla indexes intimacy and solidarity. By mixing them, users create hybrid affective positions that are legible to both peer networks and broader publics. This strategic alternation helps reconfigure existing sociolinguistic hierarchies: Bangla is no longer the default for emotional expression but is repositioned as one component in a bilingual stance repertoire.

Domain-specific innovations include hybrid compounds that combine English nouns with Bangla grammatical patterns. Examples include *study-pressure ami broken* (I am broken under study pressure), *job-tension korchhi* (I am experiencing job tension), and *family-drama shuru hoyeche* (family drama has begun). Cultural notions demand that both languages have complete semantic precision. Markers of the generation are found in artistic reinterpretations such as "*vibe korchhi*" (I am vibing), "*flex korbo*" (I will flex), "*cringe lage*" (it feels cringeworthy), and "*savage hoye gechi*" (I have become savage). These demonstrate the local incorporation of global young culture vocabulary in a grammatically correct manner.

Patterns of reverse borrowing reveal instances where terms in Bangla are borrowed into English-dominant discourses due to cultural particularity. Such words as *adda* (informal social grouping), *bhalobasha* (strong romantic love), *maya* (affection/attachment) are hard to translate because the semantic content is too complex to be encompassed by English words. A good example is the use of terms such as "We had such a good *adda* last night" or "I have so much *maya* of this place," which illustrate how culturally specific ideas can be incorporated into English discourse.

While these hybrid lexical forms are well-attested, their proliferation is not merely ornamental. We interpret their adoption as prestige-indexed stance work: younger users draw on English affect lexemes to perform cosmopolitan belonging while retaining Bangla morphosyntax for local legitimacy. This dual orientation indexes what Androutsopoulos (2015) calls "networked multilingual capital," signaling competence in global registers without disavowing national identity. This also explains why emotive and pop-culture lexemes dominate over technical jargon among younger cohorts.

Table 1 Types of lexical innovation in Banglailish.

Innovation type	Description	Examples	Share of morphosyntactic mixing instances (%)	Primary domains
Hybrid Compounds	English noun + Bangla grammatical patterns	<i>study-pressure, job-tension, family-drama</i>	23%	Academic, professional, social
Global Youth Adaptations	International slang + Bangla inflexion	<i>vibe korchī, flex korbo, cringe lage</i>	18%	Entertainment, social media
Reverse Borrowing	Bangla terms in English-dominant contexts	<i>adda, bhalobasha, maya</i>	15%	Cultural concepts, relationships
Technical Insertions	English tech terms + Bangla syntax	<i>app download kori, phone charge ache</i>	21%	Technology, digital life
Emotional Expressions	English emotion words + Bangla particles	<i>excited hoye gechi, nervous lagche</i>	12%	Personal feelings, reactions
Professional Register	Workplace English + formal Bangla	<i>meeting korbo, project complete hoyeche</i>	11%	Professional communication

Note. Percentages based on 120,000 token corpus; categories not mutually exclusive. Base = 120,000 tokens; categories not mutually exclusive.

Morphosyntactic Sophistication: Systematic Hybrid Constructions

Banglailish has morphosyntactic consistency that allows one to predict systematic variety formation. The forms of hybrid verb constructions are predictable, where English verb roots regularly combine with Bangla auxiliary verb roots and inflectional morphology: *comment korchī* (I am commenting), *share korbo* (I will share), *like koreche* (he/she has liked), *download korte hobe* (will have to download). These constructions demonstrate structural constraints, as opposed to ad hoc mixing, where certain limits are imposed on the combinations that can be made.

An example is that English main verbs readily fuse with Bangla auxiliaries (post *korchī*, chat *korbo*). In contrast, English auxiliaries cannot fuse with Bangla auxiliaries (attempts, such as “will *korbo*,” are systematically blocked). This shows the internal grammatical provisions that govern the formation of hybrids.

Noun phrase mixing reveals sophisticated adaptation strategies. Determiners and possessive markers show a strong preference for Bangla forms: *amar phone* (my phone), *ei app* (this app), *notun laptop* (new laptop), and *oder car* (their car). Content nouns in technical domains favour English insertion while maintaining Bangla grammatical frameworks.

The morphosyntactic hybridizations observed (e.g., *job-tension korch*i, *plan-ta cancel hoye geche*) appear not as random creativity but as routinized constructions. Their consistent Bangla matrix frame suggests an emergent “mixed grammar” where English stems are domesticated into Bangla slots. This indicates durability: such blends are likely to persist as semi-fixed expressions rather than ephemeral fads. The retention of Bangla auxiliaries (96%) further supports this claim, showing that even innovative structures conform to entrenched syntactic scaffolding.

Table 2 Morphosyntactic patterns in Bangllish.

Construction type	Pattern	Examples	Frequency	Constraints
Hybrid Verbs	English stem + Bangla auxiliary	<i>comment korch</i> i, <i>share korbo</i> , <i>like koreche</i>	34%	✓ English main verbs + Bangla auxiliaries; ✗ English auxiliaries + Bangla auxiliaries
Mixed Noun Phrases	Bangla determiner + English noun	<i>amar phone</i> , <i>ei app</i> , <i>notun laptop</i>	28%	Strong preference for Bangla determiners and possessives
Technical Insertions	English technical term + Bangla syntax	<i>download complete hoyeche</i> , <i>file save kora lagbe</i>	19%	Primarily in technology and professional domains
Compound Constructions	English + Bangla morphological fusion	<i>study-korte</i> , <i>job-khuja</i> , <i>exam-debo</i>	12%	Hyphenated forms are common in academic contexts
Prepositional Mixing	Mixed prepositional systems	<i>Facebook-e post korch</i> i, <i>YouTube theke dekhchi</i>	7%	Bangla postpositions preferred over English prepositions

Note. Percentages represent distribution within morphosyntactic mixing instances (*n* = 41,280 mixed constructions).

Platform Ecology: Distinct Varieties Within Bangllish

Each platform cultivates distinctive linguistic practices demonstrating users’ sophisticated awareness of audience and context. Bangladeshi Facebook discourse emphasises narrative elaboration with examples like: *Ajke office-e eto kaj chilo je lunch-o korte parini, completely exhausted hoye gechi* (I had so much work at the office today that I could not even have lunch; I became entirely exhausted). Extended discourse facilitates creative linguistic experimentation and the development of community-specific vocabulary.

Bangladeshi Twitter discourse is semantically dense and hashtag-driven, mainly due to the platform’s character restrictions. Examples include constructions with compressed word formats, such as ‘*exam stress-e dead*’ (meaning “dead due to exam stress”), and compound

hashtags, like #DhakaLife, #TechNews, and #BangladeshFirst. Political discourse shows: *Vote korbo kintu corruption bondho koro* (I will vote, but prevent corruption).

Bangladeshi YouTube discourse tends toward reaction language and incorporates English exclamations into the Bangla commentary: “*Ei video ta dekhe literally shocked hoye gechi, so funny!*” (I am literally shocked by seeing this video; it is so funny)! Users employ systematic code-mixing for emphasis: “*Outstanding performance, really proud feel korchhi*” (Outstanding performance, I am feeling really proud).

Bangladeshi Instagram discourse demonstrates the visual–textual synergy in which captions are crafted with sufficient communicative clarity and aesthetic appeal. Examples include: *Beach sunset dekhe mone hocche life is beautiful* (Being at the beach sunset makes life beautiful). Hashtag strategies are based on global search and incorporating references to local culture, including: BangladeshBeauty, DhakaVibes, and TraditionalWithModern.

Table 3 Platform-specific characteristics of Bangladeshi varieties.

Platform	Character limit	Avg. morphemes per hybrid element	Primary discourse type	Dominant features	Unique innovations
Facebook	Unlimited	4.1	Narrative elaboration, emotional expression	Extended mixed constructions, community-specific vocabulary	Creative linguistic experimentation, elaborate storytelling
Twitter	280 characters	2.3	Real-time commentary, political discourse	Semantic density, abbreviated constructions	Compound hashtags (#DhakaLife, #TechNews), compressed mixing
YouTube	10,000 characters	3.2	Reaction language, content interpretation	Emphasis and emotional expression	Global–local cultural negotiation, reaction vocabulary
Instagram	2,200 characters	2.8	Visual–textual integration, aesthetic captions	Balance of clarity and appeal	Searchable hashtags with local references, visual-linguistic harmony

Note. Data collected from 120,000 tokens across platforms (January–December, 2024). Base = 120,000 tokens; platform buckets per method.

Sociolinguistic Functions: Identity, Community, and Cultural Negotiation

Banglalish serves sophisticated sociolinguistic functions beyond effective communication. Identity performance allows users to balance global citizenship with local cultural authenticity. For example, a university student posts: *Harvard-e admission peyechi kintu Dhaka University-r memories chere jete parbo na* (I got admission to Harvard, but I will not be able to leave the memories of Dhaka University). This construction maintains international aspiration while asserting local emotional attachment.

Group solidarity manifests through generation-specific mixing patterns. Younger users (18–25) employ global youth slang with Bangla grammar: *Ei movie dekhe literally mind blown hoye gechi*, so *aesthetic* (I am literally mind blown watching this movie, so aesthetic). Older users (36–45) show more conservative patterns: *Office meeting-e presentation dite hobe* (I have to give a presentation in the office meeting).

Professional identity construction demonstrates sophisticated register variation. In formal contexts: *Project deadline extend korte parben kina janaben* (Please let me know if you can extend the project deadline). In informal contexts: *Boss amake completely frustrate kore diyeche*, *leave niye stressed* (The boss has completely frustrated me, I am stressed about taking leave). This shows highly developed metalinguistic competence in knowing when and how to use different varieties.

Cultural negotiation is evident in how users navigate global culture while preserving local authenticity: *K-pop fan hoye gechi kintu Bangla music-er proti love kome nai* (I have become a K-pop fan, but my love for Bangla music has not decreased). Such constructions demonstrate sophisticated identity management strategies.

Demographic Variations: Surprising Patterns and Unexpected Discoveries

The demographic analysis reveals patterns that both confirm and complicate previous sociolinguistic findings.

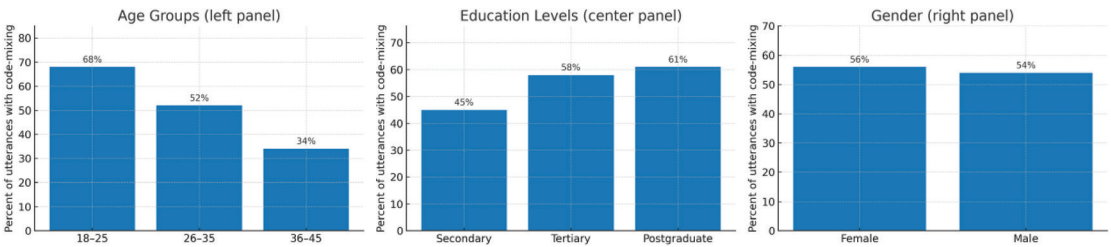


Figure 1 Demographic variations: surprising patterns and unexpected discoveries.

The chart displays the percentage of utterances containing code-mixing elements across three demographic dimensions: age groups (left panel), education levels (centre panel), and gender (right panel). Notable patterns include an inverse relationship between age and mixing frequency, a positive correlation between education level and code-mixing, and minimal gender-based differences, based on data from 120,000 tokens across Facebook, Twitter, YouTube, and Instagram (January–December, 2024).

Age-Sensitive Pattern Examples

The 18–25 age group demonstrates the highest level of innovation: *Eid shopping-e literally broke hoye gechi*, but the *outfit is on fire* (I literally went broke on Eid shopping, but the outfit is fantastic). This group shows creative emoji integration: *Exams shesh 😊 but vacation mood 😊* (Exams finished 😊 but vacation mood 😊).

The 26–35 age group shows the most systematic complexity: *Client presentation-e feedback positive chilo*, next *quarter-e expansion plan discuss korbo* (The feedback was positive in the client presentation, we will discuss the expansion plan next quarter). This group demonstrates professional register mastery.

The 36–45 age group exhibits conservative patterns: *Bacchader school admission-er jonno worried hoye achi* (I am worried about the children's school admission). Limited innovation with focus on family and traditional concerns.

Educational Level Innovations

Secondary education users show creative phonological adaptation, writing English words using Bangla phonetic patterns: *foyn* for phone, *komputer* for computer, *mesej* for message. They innovate orthographically: 'luv u' becomes 'lab u', 'thanks' becomes 'dhanksbai'.

Postgraduate users demonstrate sophisticated register distinction, switching between formal academic discourse: *Research methodology-te quantitative approach follow korchhi* (I am following a quantitative approach in research methodology) and casual conversation: *Coffee break-e adda marte jabo* (I will chat during coffee break).

Professional Register Variations

Professional contexts show systematic vocabulary clusters:

- Academic: *thesis defend korte hobe, publication submit korechi* (I have to defend the thesis, I have submitted the publication)
- Corporate: *meeting reschedule korte parben, deadline extend korte hobe* (can you reschedule the meeting, the deadline will have to be extended)
- Technical: *software update kora lagbe, bug fix complete hoyeche* (the software needs to be updated, bug fix has been completed)

Local Cultural Allusions

Users preserve local cultural references through strategic mixing:

- Religious contexts: *Eid Mubarak, celebration-e join korben?* (Eid Mubarak, will you join the celebration?)
- Cultural events: *Pohela Boishakh-e traditional dress wear korbo* (I will wear traditional dress on Pohela Boishakh)
- Local foods: *Biriyani khete mon chacche, foodie mood* (I want to eat biriyani, in foodie mood)
- Regional identity: *Chittagong-er accent miss korchhi, hometown vibes* (I miss Chittagong accent, hometown vibes)

Discussion

Our corpus indicates that everyday Bangladeshi digital discourse regularly mobilises rule-governed cross-language resources. This is evident in the 89% Bangla matrix rate in mixed clauses and the 96% selection of Bangla auxiliaries in hybrid verbs, i.e., innovation within a stable grammatical framework.

MLF Model Validation in Digital Bangla–English Code-Mixing

Corpus analysis confirms the Matrix Language Frame (MLF) model basics of prediction (Myers-Scotton, 1993, 2002) and places them within a uniquely digital ecology. In the 120,000-token sample, Bangla provides system morphemes and the clause structure by serving as the matrix language in 89% of mixed clauses. This replicates a well-attested cross-linguistic asymmetry (Muysken, 2000), creating a code-mixing typology that identified three types in which the grammatical “frame” remains stable even as content vocabulary is imported. In our material,

hybrid verb complexes such as *comment korchī*, *share korbo*, *like koreche*, and *download korte hobe* are not sporadic: they occur frequently and follow the same auxiliary sequencing as monolingual Bangla.

Two observations can be used to expand the theoretical picture. To begin with, consistency is enhanced: 96% of all hybrid verbs choose Bangla auxiliary morphology, which is moderately higher than that of most face-to-face corpora. We take this to mean a sort of effect of digital strengthening—typed, editable spaces encourage planning and self-observation, and thus more people stick to the constraints of matrix-language. Second, our data register domain clustering reveals that technology and professional registers favour English lexical stems but still integrate them within Bangla morphosyntax, reinforcing the distinction between grammatical frame and lexical import. Together, these facts affirm MLF's architecture while motivating a refinement for online writing.

Cross-Linguistic Comparison of Insertional Patterns

Patterns observed in Banglailish align with insertional mixing reported in Welsh-English (Deuchar, 2006), Turkish-German (Backus, 1996; Çetinoğlu, 2016), and Spanish-English (Callahan, 2002): one language reliably provides the morphosyntactic frame, while the other contributes content morphemes. In our dataset, this manifests as Bangla determiners, postpositions, and auxiliaries that host English nouns and verb stems. The result strengthens claims that matrix-language dominance is not limited to particular language pairs but generalises to new, digitally mediated contexts.

However, a purely structural comparison is incomplete without attending to platform affordances. As summarized in Table 3, Facebook encourages extended mixed turns (averaging 4.1 morphemes per hybrid element), whereas Twitter promotes compressed, hashtag-dense forms (2.3). YouTube threads concentrate evaluative bursts, and Instagram captions orchestrate visual-text couplings. Adopting Androutsopoulos's (2015) *networked multilingualism* and Lexander and Androutsopoulos's (2023) *medial repertoires*, we argue that the same speakers redistribute a stable grammatical repertoire across platforms to meet audience design, visibility, and interactional needs. Thus, cross-linguistic similarities reside at the level of structure, while cross-platform differences play out at the level of packaging and performance.

Morphosyntactic Sophistication and Novel Constructions

The morphosyntactic regularities found in Bengali are confirmed and extended in Bengali by Banglailish. Table 2 suggests a strong tendency for mixed noun phrases to use Bengali determiners and possessives (*ei app*, *amar phone*, *notun laptop*). By contrast, English content nouns are more likely to be concentrated in the field of technology and professional worlds. No less systematic are hybrid verb complexes that import Bangla auxiliary morphology to accommodate English stems (*post korchī*, *chat korbo*), with certain constraints apparent: English auxiliaries cannot stack with Bangla auxiliaries, and constructions such as *will korbo* are frowned upon.

In addition to these canonical configurations, the dataset also exhibits productive compounding, which puts simple insertion accounts under strain. In Table 1, lexical innovations are explained by the hybrid compounds, including the study-pressure, job-tension, and family-drama, which represent 23% of the innovations. These are semantically clear but morphologically inventive, indexing stance and affect without being incongruent with the grammar of Bangla. Translanguaging makes this understandable: they use a unified, combined semiotic repertoire (García & Li, 2014; Li, 2018), which is why global youth slang (*vibe korchī flex korbo cringe lage*) does not seem like switching to Bangla morphology. Creativity is not a flight out of rule-regulated behaviour; it is creativity within a stable structural architecture.

Digital Platform Effects and Medial Repertoires

We argue that platform affordances repackage a stable grammatical repertoire (Bangla matrix + English content morphemes) into distinct performance formats, rather than producing entirely new grammars. Facebook permits the elaboration of narrative mixing through the use of community-specialised vocabulary; Twitter diminishes stance/evaluation/reference to small hybrids and compound hashtags; YouTube accommodates short evaluative formulations interspersed with English intensifiers; Instagram aligns captions and visual appearance with searchable labels. A comparable affordance-driven reconfiguration is reported in the recent “TikTok-refugee” migration to Xiaohongshu (RedNote), where notebook-style posts and searchable tags scaffolded informal translingual learning practices (see *TikTok Refugees, Digital Migration, and the Expanding Affordances of Xiaohongshu*, 2025). These comparisons, reported in Table 3, relate to various types of activities (storytelling, commentary, reaction, curation) and different attention economies (algorithmic visibility, scannability, searchability). Because affordances condition packaging, the same Bangla-matrix repertoire lengthens on Facebook (avg. 4.1 morphemes per hybrid element) but compresses on Twitter (down to 2.3), while Instagram orchestrates text-visual pairings.

Following *networked multilingualism* (Androutsopoulos, 2015) and *medial repertoires* (Lexander & Androutsopoulos, 2023), we interpret cross-platform differences as repertoire reconfiguration. Speakers mobilise the same grammatical resources, Bangla matrix morphology with embedded English stems, but package them with emojis, mentions, and hashtags to optimise uptake. The MLF model still locates the grammatical “where” and “how”; platform theories explain the “why” and “for whom.” In this sense, grammar is necessary but not sufficient for predictive adequacy in digital settings: distribution, uptake, and participation frameworks complete the explanatory arc.

Demographic Patterns and Identity Work

Age and education correlate with both frequency and complexity of mixing. Younger users (18–25) post the densest innovation, often aligning English youth slang with Bangla grammar and emoji-rich stance designs (e.g., “*Exams shesh* 😊 but vacation mood 😊”). Users with postgraduate training show register-sensitive mixing, alternating between professional tasks (“*deadline extend korte hobe*”) and casual chat (“*adda marte jabo*”). Minimal gender differences suggest that digital publics dampen some traditional constraints (Herring & Stoerger, 2014), consistent with superdiversity accounts that emphasise fluid social alignments online (Varis & Wang, 2015; Vertovec, 2007).

More importantly, these demographic trends cannot be attributed solely to deficiency or partial acquisition. They can be interpreted through a translanguaging prism (García & Li, 2014; Li, 2018): speakers who are younger index cosmopolitan belonging; those in professional settings negotiate work demands without losing their local voice; those older use more conservative and less hybridised forms. The Bangla matrix is always constant—it is the density, indexicality, and multimodal elaboration of the repertoire that changes in each instance.

Challenges to Traditional MLF Accounts

Classical assumptions are under pressure from two families of findings. First, reverse borrowing, i.e., the application of culturally dense Bangla words (*adda*, *bhalobasha*, *maya*) within otherwise English-dominated stretches, creates two-way lexical movement and is not well explicable in terms of prestige. Second, world youth-culture products, when incorporated into Bengali morphology (e.g., “*savage hoye gechi*,” “*cringe lage*”), are challenging to define as either a borrowing or a loan. These phenomena are congruent with what Varis and Wang (2015) define

as superdiverse repertoires, where boundaries are blurred across languages, modalities, and styles.

Instead of conflicting with the MLF model, these patterns expose the scope conditions of the model. MLF sparsely represents grammatical structure; it is not explicitly an indexical meaning architecture or platform-conditioned uptake. On augmenting with translanguaging and networked multilingualism, we get a layered account: a stable matrix determines clause building, a flexible repertoire determines social meaning, cross-platform portability, and resourcefulness in innovation.

Theoretical and Educational Implications

In theory, our data favour a synthesis: (i) MLF describes the structure of Banglailish-labelled as matrix dominance, system morphemes, and the restriction of hybrid verb complexes, (ii) translanguaging explains the mobilisation of mixed resources—stance, identity, alignment, and creative expressivity by speakers, (iii) networked multilingualism and medial repertoires describe the redistribution of the same grammatical kit among activity types, interactional rhythms, and visibility regimes by platform affordances, and (iv) superdiversity describes how new lexical and multimodal conventions stabilise in digitally connected publics (scale and speed).

Our results generalise previous theoretical constructs in several ways. On the one hand, they show that the MLF model can be kept in digital writing with some modifications to capture the effects of digital strengthening that enhance the consistency of the matrix language in typed setups. Second, they demonstrate that translanguaging should be expanded to incorporate multimodal semiotic tools like emojis, hashtags, and visual-text couplings that are in the focus of the meaning-making in Banglailish. Third, they promote the idea of networked multilingualism, demonstrating that platform-specific affordances repackage the identical grammatical repertoire to be relevant in different interactional contexts. All these insights are a step toward developing multilingual theory in digitally mediated settings.

Practically, the findings argue for rethinking language education and assessment in Bangladesh and similar postcolonial contexts. We outline four actionable domains:

1. **Curriculum design:** integrate genre-authentic digital tasks, caption writing, threaded debates, reaction summaries, and multimodal storytelling that invite strategic mixing. Students should be taught to justify mixing choices in terms of audience, purpose, and platform affordances rather than penalised for “impurity.” Caption rewrite across platforms (Facebook → Twitter) with a brief rationale for mixing choices due to different affordances.
2. **Assessment:** include rubrics for audience design, multimodal orchestration (text + emoji + tag), and register management (e.g., shifting between workplace and informal discourse) alongside grammatical accuracy. This treats **digital multilingual competence** as a 21st-century literacy rather than a deviation from standard norms. Rubric adds “emoji/tag orchestration” and “audience design” criteria alongside grammatical accuracy.
3. **Teacher professional development:** provide modules on translanguaging pedagogy, corpus-informed platform discourse, and feedback that targets communicative effectiveness (clarity, stance, tone, ethical citation) in mixed varieties. Classroom forums and LMS platforms offer low-stakes venues for guided practice. Workshop on giving feedback for clarity, stance, and register in mixed varieties using real corpus snippets.
4. **Policy and standards:** national curricula and QA frameworks should explicitly recognise mixed digital genres as legitimate learning outcomes. Institutional policies can

encourage bilingual signage, mixed-genre academic assignments, and research dissemination in locally meaningful hybrid registers. Permit mixed-genre assignments (e.g., bilingual abstracts; hashtagged summaries) in institutional guidelines.

Lastly, the research also initiates three research paths. Longitudinal study designs can be used to test the hypothesis that the specific hybrids (e.g., *comment korchī*) stabilise or cycle out due to changing platform incentives. The cost and cue of processing of reading/writing hybrid verbs and mixed noun phrases can be tested using psycholinguistic experiments. International comparisons of corpora (e.g., South Asia, the Middle East, and North Africa, or MENA) can be categorised as either universal architecture (matrix dominance) or situation-specific stylistic repertoires. These agendas receive methodological scaffolding in recent programmatic work on social-media corpora (Rüdiger & Dayter, 2020) and variation in world Englishes (Schneider, 2020).

Conclusion: Toward Multilingual Digital Futures

This corpus-based research into the digital discourse in Bangladesh evidences that social media networks are not places of disorganized language contact but of ordered linguistic invention. The 120,000-token sample provides evidence of rule-determined trends of Bangla-English integration that undermine deficit models and support structural constraints as hypothesized by the contact linguistics theory.

Empirical Contributions

The research paper comes up with three main empirical results. To begin with, Banglali has regular grammatical patterns with Bangla as the matrix language in 89% of mixed clauses, as well as the morphology of auxiliaries in 96% of hybrid verb structures. Such percentages reflect structural stability, not transitional mixing. However, longitudinal data would be required to determine whether these trends are permanent varieties formation or adaptations to digital realms at the current time.

Second, platform affordances condition in a systematic way the deployment of multilingual repertoires by users. The cross-platform difference observed here between the extended narrative combining (4.1 morphemes per hybrid element) of Facebook and the parsimonious (2.3 morphemes) of Twitter is evidence that speakers are actively reforming the stable grammatical resources to suit various communicative needs. This result extends the networked multilingualism theory by providing quantitative data on platform-specific adaptation strategies.

Third, demographic trends indicate that there is advanced register awareness and not language loss by the generation. The choice of English by younger users influences terms (*excited hoye gechi* versus *uttejito hoye gechi*), or strategies of identity positioning in the global youth discourse, and not the lack of bilingual competence. Nevertheless, the 12-month period does not allow for concluding on the need for such preference stabilizing as community norms or changing with the dynamics of cultural affiliations.

Theoretical Implications

The results support a layered theoretical framework that combines structural and poststructural views. The Frame model of Matrix Language precisely forecasts linguistic limits in the digital environment, and the digital environment has the potential to improve consistency in the matrix language through the plan affordances of written communication. At the same time, the theory of translanguaging refers to the artistic mobilization of language resources towards identity work and community building.

This summary indicates that structural models continue to be needed to describe the grammatical structure of contact varieties, and sociolinguistic theories account for the social reasons behind the innovation. Nevertheless, the study lacks breadth, as it is conducted among urban, educated, and well-connected internet users, which limits the transferability of the results to the overall linguistic ecology of Bangladesh.

Methodological Limitations and Future Directions

There are several constraints to the study's claims. The emphasis on social posts is a lockout of private messaging activities that can reveal various mixing patterns. The 12-month period of collection, though covering the political changes, cannot conclude whether the observed tendencies are a stable evolution or a short-term adjustment. The geographic limitation of cities cannot take away the possibility of exploring the rural digital practices, yet this is the reality of the current internet penetration.

There are three directions that future research should take. Keeping track of the same users over several years and longitudinal studies would help distinguish between stable and temporary trends in innovations. Cognitive models of the production of bilinguals in digital situations could be tested in psycholinguistic experiments testing the processing costs of hybrid constructions. Comparisons with other postcolonial digital multilingual communities in other countries would help to determine universal and culture-specific patterns of language contact mediated by the platform.

Educational and Policy Considerations

These attested levels of sophistication of multilingual practices in the digital environment indicate the existence of the need to establish pedagogical frameworks that would not be stigmatizing of strategic code-mixing. Schools may establish assessment standards that assess communicative competence in multilingual repertoires and still uphold the academic standards. Nonetheless, it would need to be implemented with sensitivity to socioeconomic differences in access to digital resources and English language proficiency.

The implications of language policy are not known because the study is limited. Although the results indicate innovation among young, educated urban citizens, policy decisions that impact the national education or official language would need more demographic representation and further study of the attitudes across the social strata.

Concluding Observations

The appearance of systematic trends in Bangladeshi online discourse is not a symptom of the decline of Bangla or the dominance of English, but rather the formation of new multilingual competencies adjusted to the new communicative reality. Users demonstrate advanced audience, context, and platform awareness, and exhibit grammatical consistency across multiple digital contexts. It is still a question of future research that needs to be followed to establish whether those practices imply a long-term shift of the language or a short-term technological adjustment. The trends recorded herein show that postcolonial multilingual societies are in the process of redefining inherited language hierarchies using digital innovation. However, it is not possible to forecast the long-term effects of these changes based on the existing evidence.

This work adds to the discussion on how communication mediated by platforms affects the outcomes of language contact and supports the necessity of a more comprehensive, longitudinal study to find out whether the digital multilingual practices are transitional responses to the environment or permanent changes in the language ecologies of the postcolonial period.

Ethical Statement

In this research, the data gathered (socially available publicly) was analysed and thus did not involve the Institutional Review Board (IRB) approval. All data were anonymised entirely, and the study was conducted in accordance with internationally accepted ethical standards for online data gathering and analysis, including those suggested by the Association of Internet Researchers (AoIR).

Declaration of Conflicting Interests

The author declares no conflict of interest. No personal, financial, or institutional interests influenced the design, analysis, or reporting of this study.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted entirely through the author's independent efforts and resources, without external financial support or institutional backing.

References

- Al Jazeera. (2024, August 8). *Muhammad Yunus takes oath as head of Bangladesh's interim government*. <https://www.aljazeera.com/news/2024/8/8/muhammad-yunus-takes-oath-as-head-of-bangladeshs-interim-government>
- Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19(2), 185–205. <https://doi.org/10.1177/1367006913489198>
- Backus, A. (1996). *Two in one: Bilingual speech of Turkish immigrants in the Netherlands*. Tilburg University Press.
- Backus, A. (2003). Units in code-switching: Evidence for multimorphemic elements in the lexicon. *Linguistics*, 41(1), 83–132. <https://doi.org/10.1515/ling.2003.005>
- Blommaert, J. (2010). *The sociolinguistics of globalization*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511845307>
- Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Callahan, L. (2002). The Matrix Language Frame model and Spanish/English codeswitching in fiction. *Language & Communication*, 22(1), 1–16. [https://doi.org/10.1016/S0271-5309\(01\)00018-0](https://doi.org/10.1016/S0271-5309(01)00018-0)
- Canagarajah, S. (2012). *Translingual practice: Global Englishes and cosmopolitan relations*. Routledge. <https://doi.org/10.4324/9780203073889>
- Çetinoğlu, Ö. (2016). A Turkish–German code-switching corpus. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* (pp. 4215–4220). ELRA.
- Deuchar, M. (2006). Welsh–English code-switching and the Matrix Language Frame model. *Lingua*, 116(11), 1986–2011. <https://doi.org/10.1016/j.lingua.2004.10.001>
- Deumert, A. (2014). *Sociolinguistics and mobile communication*. Edinburgh University Press. <https://doi.org/10.1515/9780748655755>
- Dizon, G., Gold, J., & Barnes, R. (2025). ChatGPT for self-regulated language learning: University English as a foreign language students' practices and perceptions. *Digital Applied Linguistics*, 3, Article 102510. <https://doi.org/10.29140/dal.v3.102510>
- García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Wiley-Blackwell.
- García, O., & Li, W. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan. <https://doi.org/10.1057/9781137385765>
- Gumperz, J. J. (1982). *Discourse strategies (No. 1)*. Cambridge University Press.
- Hamid, O. M., Sultana, S., & Roshid, M. M. (2025) Transformation of applied linguistics in the global south context of Bangladesh: Researcher agency, imagination, and north-south cooperation. *Applied Linguistics*, 46(4), 557–575. <https://doi.org/10.1093/applin/amae037>

- Herring, S. C., & Stoerger, S. (2014). Gender and (a)nonymity in computer-mediated communication. In S. Ehrlich, M. Meyerhoff, & J. Holmes (Eds.), *The handbook of language, gender, and sexuality* (2nd ed., pp. 567–586). Wiley-Blackwell. <https://doi.org/10.1002/9781118584248.ch29>
- Huang, J., & Mizumoto, A. (2024). Examining the effect of generative AI on students' motivation and writing self-efficacy. *Digital Applied Linguistics*, 1, Article 102324. <https://doi.org/10.29140/dal.v1.102324>
- Huang, J., & Teng, M. F. (2025). Peer feedback and ChatGPT-generated feedback on Japanese EFL students' engagement in a foreign language writing context. *Digital Applied Linguistics*, 2, Article 102469. <https://doi.org/10.29140/dal.v2.102469>
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11–30). Cambridge University Press.
- Lee, C. (2017). Multilingualism online. In S. Seargeant & C. Tagg (Eds.), *The Routledge handbook of language and digital communication* (pp. 403–418). Routledge.
- Lexander, K. V., & Androutsopoulos, J. (2023). *Multilingual families in a digital age: Medial repertoires and transnational practices*. Routledge. <https://doi.org/10.4324/9781003227311>
- Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, 39(1), 9–30. <https://doi.org/10.1093/applin/amx039>
- Liu, G. L., Ng, K. H., & Gonzales, W. D. W. (2024). The predictive role of L2 motivation in receptive and productive informal digital learning of English: A chain mediation model. *Digital Applied Linguistics*, 1, Article 2229. <https://doi.org/10.29140/dal.v1.2229>
- Ma, Y., & Li, W. (2024). Unpacking translanguaging in refusals on Chinese social media: Strategies, distribution, and functions. *Multilingua*, 43(4), 487–524. <https://doi.org/10.1515/multi-2023-0071>
- Mahboob, A. (2014). Understanding language variation: Implications for EIL pedagogy. In R. Marlina & R. A. Giri (Eds.), *The pedagogy of English as an international language* (pp. 257–268). Springer.
- Makoni, S., & Pennycook, A. (2006). *Disinventing and reconstituting languages*. Multilingual Matters. <https://doi.org/10.21832/9781853599255>
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511486289>
- Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford University Press.
- Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford University Press.
- Myers-Scotton, C., & Jake, J. L. (2017). Revisiting the 4-M model and Matrix Language Frame model: English as a morpheme language and language level. *Bilingualism: Language and Cognition*, 20(4), 681–692.
- Norton, B. (2013). *Identity and language learning: Extending the conversation* (2nd ed.). Multilingual Matters.
- Pennycook, A. (2017). *The cultural politics of English as an international language* (1st ed.). Routledge. <https://doi.org/10.4324/9781315225593>
- Poplack, S. (1980). Sometimes I will start a sentence in Spanish and terminate it in Spanish: Toward a typology of code-switching. *Linguistics*, 18(7–8), 581–618. <https://doi.org/10.1515/ling.1980.18.7-8.581>
- Rüdiger, S., & Dayter, D. (2020). Introduction: The expanding landscape of corpus-based studies of social media language. In S. Rüdiger & D. Dayter (Eds.), *Corpus approaches to social media* (pp. 1–12). John Benjamins. <https://doi.org/10.1075/sci.98.int>
- Rymes, B., & Leone, A. W. (2014). Citizen sociolinguistics: A new media methodology for understanding language and social life. *Working Papers in Educational Linguistics*, 29(2), 25–43.
- Schneider, E. W. (2020). *English around the world: An introduction* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/9781108656726>
- Seargeant, P., & Tagg, C. (2014). Introduction: The language of social media. In P. Seargeant & C. Tagg (Eds.), *The language of social media* (pp. 1–20). Palgrave Macmillan. https://doi.org/10.1057/9781137029317_1
- Tagg, C. (2015). *Exploring digital communication: Language in action*. Routledge. <https://doi.org/10.4324/9781315727165>

- Tao, L., Zhao, J., & Dong, J. (2024). A multimodal analysis of the online translanguaging practices of international students studying Chinese in a Chinese university. *Applied Linguistics Review*, 15(4), 1531–1555. <https://doi.org/10.1515/applirev-2022-0139>
- Thurlow, C., & Mroczek, K. (2011). *Digital discourse: Language in the new media*. Oxford University Press.
- Varis, P., & Wang, X. (2015). Superdiversity on the Internet: A case from China. *Diversities*, 13(2), 71–83.
- Vertovec, S. (2007). Super-diversity and its implications. *Ethnic and Racial Studies*, 30(6), 1024–1054. <https://doi.org/10.1080/01419870701599465>