

SOCIOLINGUISTIC ANALYSIS OF CODE-MIXING PRACTICES IN INDONESIAN BEAUTY VLOGGERS' YOUTUBE CONTENT: TYPES AND FUNCTIONS

Sri Dwi Lestari Bobihu¹, Sarah Revadiah Panigoro², Adriansyah Abu Katili³

lestaribobihu07@gmail.com¹, revadiahpanigoro@gmail.com², adriansyahkatili@ung.ac.id³

Universitas Negeri Gorontalo

ABSTRACT

Code-mixing, the simultaneous use of two or more languages within a single discourse, has become a pervasive feature of digital communication in Indonesia. This study examines code-mixing practices among popular Indonesian beauty vloggers on YouTube, a platform that commands millions of viewers and functions as a significant site of informal language socialisation. Drawing on Muysken's (2000) typological framework—which distinguishes between insertion, alternation, and congruent lexicalisation—and integrating sociolinguistic and psycholinguistic functions identified by scholars such as Hoffman (1991) and Auer (1998), this paper offers a qualitative descriptive analysis of code-mixing patterns documented across ten purposively selected video transcripts. Findings reveal that insertion of English lexical items into Indonesian utterances constitutes the dominant code-mixing type, frequently serving functions of lexical gap-filling, identity negotiation, and audience rapport-building. Alternation and congruent lexicalisation, while less frequent, emerge in contexts requiring affective emphasis or signalling cosmopolitan expertise. The study contributes to a growing body of research on language change in Indonesian digital spaces and carries implications for language educators navigating increasingly hybrid communicative norms among Indonesian youth.

Keywords: Code-Mixing, Indonesian, Youtube, Beauty Vloggers, Muysken, Digital Media, Sociolinguistics.

INTRODUCTION

Indonesia is one of the world's most linguistically complex nations, home to over 700 regional languages alongside its national lingua franca, Bahasa Indonesia, and a rapidly expanding sphere of English use. In the past decade, the proliferation of digital platforms—YouTube, TikTok, Instagram, and podcast services—has created a new communicative arena where Indonesian speakers routinely navigate multiple languages within a single interaction. Among the most visible of these bilingual speakers are beauty content creators, or “beauty vloggers”, whose YouTube channels attract audiences ranging from tens of thousands to tens of millions of subscribers. These creators consistently blend Indonesian and English in ways that feel effortless, stylish, and distinctly contemporary, making them compelling subjects for sociolinguistic inquiry.

The phenomenon they engage in—code-mixing—refers to the embedding of elements from one language into the grammatical structure of another (Muysken, 2000). While related to code-switching, which typically involves shifting between languages at sentence boundaries, code-mixing occurs within a single utterance or clause, and is thus a more intimate form of bilingual performance. The literature distinguishes code-mixing from code-switching primarily on structural grounds, though in everyday analysis the terms are sometimes used interchangeably (Poplack, 1980; Wardhaugh & Fuller, 2015). For the purposes of this study, code-mixing is defined following Muysken's (2000) framework, which provides a systematic typological account.

Despite the abundance of sociolinguistic research on code-switching and code-mixing in Indonesian contexts particularly in classroom discourse, print media, and online forums there remains a notable gap in the literature concerning code-mixing as it manifests in video-based digital content. The few studies that have begun to address this gap (Auliabella, 2022; Fauziah & Suryani, 2022) focus on individual vloggers, leaving questions about cross-channel patterns and function largely unresolved. Beauty vloggers represent a particularly significant niche: they are aspirational figures who mediate global beauty culture for local audiences, and their language use reflects and reinforces emerging bilingual norms among Indonesian youth. Understanding how and why they mix codes is therefore not merely a descriptive exercise; it carries implications for identity formation, cultural negotiation, and language education policy in contemporary Indonesia.

This study therefore aims to: (1) identify and classify the types of code-mixing present in Indonesian beauty vloggers' YouTube content using Muysken's (2000) typological framework; (2) analyse the sociolinguistic and psycholinguistic functions that motivate these code-mixing practices; and (3) discuss the implications of these practices for understanding language change in Indonesian digital media. The study is driven by the conviction that digital communicative practices warrant serious linguistic attention, particularly as they increasingly shape how younger Indonesians conceptualise language, identity, and expertise.

RESEARCH METHODS

This study adopts a qualitative descriptive approach, consistent with the interpretive tradition in sociolinguistics (Mesthrie et al., 2009). The primary data consist of transcripts drawn from ten YouTube videos produced by five established Indonesian beauty vloggers selected through purposive sampling based on subscriber count (minimum 500,000 subscribers), content focus (makeup tutorials and product reviews), and linguistic profile (documented Indonesian-English bilingual use). The channels include both macro-influencers with national reach and mid-tier creators known for high audience engagement. Video selection prioritised diversity of content type: first-impression reviews, routine tutorials, and opinion-based vlogs.

All videos were manually transcribed by the researcher and a trained transcription assistant, yielding approximately 18,000 words of data. Code-mixed instances were identified through line-by-line close reading, defined operationally as any point at which an utterance in Bahasa Indonesia incorporated one or more elements from English (or vice versa) within the same clause or sentence. Each instance was then classified according to Muysken's (2000) three-way typology: insertion, alternation, or congruent lexicalisation. To ensure analytical reliability, a 20% sample of coded instances was independently reviewed by a second researcher; intercoder agreement reached 87%, which is considered acceptable for qualitative linguistic analysis. The functions of code-mixing were interpreted through the motivational framework proposed by Hoffman (1991), supplemented by empirical observations from recent Indonesian-context studies (Auliabella, 2022; Fauziah & Suryani, 2022; Situmorang et al., 2023).

HASIL DAN PEMBAHASAN

1. Overview of Code-Mixing Types

Across the ten transcripts analysed, a total of 312 code-mixed instances were identified. Of these, 214 (68.6%) were classified as insertion, 64 (20.5%) as alternation, and 34 (10.9%) as congruent lexicalisation. The predominance of insertion is consistent with prior studies of Indonesian-English contact in digital content (Auliabella, 2022; Qisthi et al.,

2022) and reflects the structural asymmetry between the two languages: Indonesian functions as the primary grammatical scaffolding, into which English lexical items particularly nouns, verbs, and adjectives from the domain of beauty and digital culture are embedded. Table 1 presents a classification of code-mixing types with representative examples drawn from the corpus.

Table 1. Classification of Code-Mixing Types in Indonesian Beauty Vlogger Content

Type	Example (Utterance)	Mixed Element	Function
Insertion	<i>"Jadi kalian bisa pakai shade ini kalau mau achieve that dewy finish, ya."</i>	achieve that dewy finish (English VP)	Lexical gap-filling; Prestige marking
Insertion	<i>"Foundation ini beneran full coverage banget, literally ga kelihatan pore-nya."</i>	literally, pore (English adverb + noun)	Emphasis; Technical precision
Insertion	<i>"Kita bakal unboxing dulu produk-produknya sebelum review."</i>	unboxing, review (English verbs)	Lexical borrowing; Digital register
Alternation	<i>"Kalau menurutku sih bagus. But honestly, I feel like it oxidises too fast on my skin."</i>	But honestly... (full English clause)	Affective shift; Audience intimacy
Alternation	<i>"Ini untuk yang kulit berminyak. If you have dry skin, this is perfect for you."</i>	Complete English sentence	Audience targeting; Inclusivity
Congruent Lexicalisation	<i>"Shade-nya match banget sama undertone aku yang warm."</i>	Blend of Ind. grammar + Eng. lexis	Cosmopolitan identity construction
Congruent Lexicalisation	<i>"Teksturnya super lightweight dan non-sticky di kulit."</i>	Hybrid morphosyntax	Technical credibility; Brand authenticity

2. Insertion: Filling Gaps and Building Prestige

As Table 1 illustrates, insertion is by far the most prevalent code-mixing strategy in the corpus. The most common pattern involves the embedding of English content-area vocabulary terms like coverage, shade, undertone, dewy, and pore into otherwise Indonesian utterances. In example (1), "achieve that dewy finish" is inserted into an Indonesian clause as a verb phrase, functioning as a kind of technical idiom for which no equivalent Indonesian phrase carries the same connotational weight in the beauty community. This aligns with Hoffman's (1991) lexical gap-filling function: speakers borrow from English not necessarily

because they lack the Indonesian word, but because the English term carries greater specificity, cultural resonance, or market legitimacy within the beauty register. Auliabella (2022) documented a comparable pattern in her analysis of Molita Lin's beauty tutorials, where English beauty terminology was inserted consistently to mark technical expertise and credibility.

A second insertion pattern involves the use of high-frequency English adverbials and discourse markers literally, honestly, basically, so to intensify or frame utterances. Example (2), “literally ga kelihatan pore-nya,” exemplifies this. The adverb “literally,” embedded as a sentence-initial modifier, performs an emphatic function that its Indonesian counterpart “secara harfiah” cannot replicate with the same colloquial force. This pattern mirrors findings from Istifci (2016) on social media code-mixing, where English discourse markers serve as affective intensifiers rather than logically precise operators. Their adoption by vloggers suggests that English is increasingly indexical of a youthful, informal register—a semiotic resource rather than merely a linguistic one consistent with the identity-marking function documented by Yuliana et al. (2015) among Indonesian public figures.

The third sub-pattern of insertion borrowing digital and commercial English verbs such as unboxing, review, and collab (short for collaboration) reflects the penetration of Anglo-American digital culture into Indonesian internet discourse. These words, already established in global YouTube culture, are inserted and often morphologically adapted to Indonesian grammar (e.g., “me-review,” “di-unboxing”), demonstrating a productive interface between borrowing and nativisation (Kridalaksana, 2008). Their use signals membership in a transnational community of beauty content creators a finding that aligns with Fauziah and Suryani's (2022) observation that Indonesian beauty vloggers strategically employ English vocabulary to project cosmopolitan credibility and professional-grade expertise.

3. Alternation: Shifting Register and Addressing Multiple Audiences

Alternation, though less frequent, is notable for the pragmatic sophistication it reveals. As shown in examples (4) and (5), alternation typically involves a shift to a complete English clause or sentence following an Indonesian one. This pattern is particularly common when creators address product suitability for specific skin types or tones a moment when they consciously reach out to an imagined subset of their audience who may be more comfortable with English. This audience-accommodation function is consistent with the findings of Auliabella (2022), who observed similar alternation patterns in beauty tutorial contexts, noting that vloggers appear to toggle between languages as a way of extending the reach of their message across a linguistically diverse viewership.

In example (4), the transition “But honestly, I feel like it oxidises too fast on my skin” is significant for its affective tenor. The shift to English here marks a moment of personal disclosure—a candid critique that the creator perhaps feels more comfortable delivering in English, where the tone is somewhat softened by distance from the primary language of the video. This phenomenon, discussed by Nilep (2006) as “the affective switch,” suggests that for bilingual speakers, one language may carry different emotional valence than another, and shifts between them can be a means of modulating self-presentation. Situmorang et al. (2023) observed a related dynamic in podcast discourse, where speakers' code-switching to English coincided with moments of self-disclosure and opinion-giving, suggesting that this pattern is not unique to the beauty vlogging genre but reflects a broader affordance of Indonesian-English bilingualism in digital communication.

4. Congruent Lexicalisation: Hybrid Expertise and Identity Construction

The least frequent but arguably most linguistically sophisticated type is congruent lexicalisation, exemplified in instances (6) and (7). Here, the grammatical structures of

Indonesian and English are sufficiently aligned that the speaker draws lexical items from both simultaneously, producing utterances that cannot be cleanly assigned to either language. The phrase “Shade-nya match banget sama undertone aku yang warm” is a case in point: the nominal suffix “-nya” and the intensifier “banget” are Indonesian, while “shade,” “match,” “undertone,” and “warm” are English beauty-register terms. The result is not a hybrid of two complete systems but rather a single, integrated system drawing on a shared bilingual lexical pool precisely what Muysken (2000) describes as the hallmark of congruent lexicalisation. Qisthi et al. (2022) similarly documented this blending pattern on Devina Hermawan's channel, arguing that it reflects a degree of bilingual integration where the two codes have become functionally inseparable in certain registers.

This type of code-mixing is analytically significant because it suggests that for some Indonesian beauty vloggers, the English-Indonesian bilingual system functions as a single, internally coherent communicative resource rather than two separate codes. This has implications for identity: congruent lexicalisation projects expertise (the speaker operates fluently at the interface of both languages) while also performing what Boztepe (2003) terms “cosmopolitan authenticity” the idea that one's credibility as a beauty commentator is partly constituted by demonstrating fluency in the global language of beauty culture. As Yuliana et al. (2015) argued in their comparative study of Indonesian celebrities, this kind of hybrid linguistic performance is not incidental but constitutive of a public persona built around bilingual sophistication. This construction of credibility through language choice is particularly potent in a digital economy where perceived expertise directly influences audience trust and monetisation.

CONCLUSION

This study has demonstrated that code-mixing in Indonesian beauty vloggers' YouTube content is a structured, multifunctional communicative practice, not a random performance of incomplete bilingualism. Drawing on Muysken's (2000) typological framework, the analysis identified insertion as the dominant pattern, accounting for over two-thirds of all code-mixed instances, followed by alternation and congruent lexicalisation. Each type is associated with distinct communicative functions: insertion primarily serves lexical precision and prestige marking; alternation facilitates audience accommodation and affective modulation; and congruent lexicalisation constructs a sophisticated cosmopolitan expertise identity. These findings are consistent with and extend existing research on code-mixing in Indonesian digital discourse (Auliabella, 2022; Fauziah & Suryani, 2022; Qisthi et al., 2022; Yuliana et al., 2015), and they confirm the applicability of Muysken's framework to the evolving landscape of Indonesian video-based digital content.

The implications of these findings are threefold. First, for sociolinguists, this study adds to accumulating evidence that digital media platforms are transformative sites of language change, where informal norms of bilingual use are being rapidly established and naturalised for mass audiences. Second, for language educators, the prevalence and systematicity of code-mixing among influential content creators suggests a need to engage rather than dismiss hybrid language practices in pedagogical contexts. Students of English are encountering and producing code-mixed language constantly; understanding its structure and functions may actually enhance metalinguistic awareness and communicative flexibility. Third, for media and cultural scholars, the findings underscore the role of beauty vloggers as agents of language socialisation who shape the bilingual norms of millions of young Indonesians, often more powerfully than formal institutional contexts.

Future research would benefit from expanding the data corpus to include other digital genres gaming livestreams, educational TikTok content, and Indonesian podcasts to test

whether the patterns identified here generalise across digital registers, building on the preliminary evidence from Situmorang et al.'s (2023) podcast study. Longitudinal research tracking shifts in code-mixing frequency and type over time could also illuminate broader trajectories of language change in digital Indonesia. Finally, audience-side studies investigating how viewers perceive and respond to code-mixed language would provide a fuller picture of the sociolinguistic ecology in which these practices are embedded.

REFERENCES

- Auer, P. (1998). *Code-switching in conversation: Language, interaction and identity*. Routledge.
- Auliabella, P. P. D. (2022). The use of Indonesian-English code mixing in one of Molita Lin's YouTube video as beauty vlogger. *Prosodi*, 16(2), 109–119.
- Boztepe, E. (2003). Issues in code-switching: Competing theories and models. *Working Papers in TESOL & Applied Linguistics*, 3(2), 1–27.
- Cahyani, H., de Courcy, M., & Barnett, J. (2018). Teachers' code-switching in bilingual classrooms: Exploring pedagogical and sociocultural functions. *International Journal of Bilingual Education and Bilingualism*, 21(4), 465–479.
- Fauziah, N., & Suryani, L. (2022). Code mixing in Indonesian beauty vloggers. *PROJECT (Professional Journal of English Education)*, 5(3), 595–604.
- Hoffman, C. (1991). *An introduction to bilingualism*. Longman.
- Istifci, I. (2016). A sociolinguistic analysis of the motivations for code-mixing in social media. *Journal of Language and Linguistic Studies*, 12(1), 1–13.
- Kridalaksana, H. (2008). *Kamus linguistik* (4th ed.). Gramedia Pustaka Utama.
- Mesthrie, R., Swann, J., Deumert, A., & Leap, W. L. (2009). *Introducing sociolinguistics* (2nd ed.). Edinburgh University Press.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge University Press.
- Nilep, C. (2006). “Code switching” in sociocultural linguistics. *Colorado Research in Linguistics*, 19, 1–22.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, 18(7–8), 581–618.
- Qisthi, A. F., Ricahyono, S., & Chasanatun, T. W. (2022). Global culture, cultural hybridity, and intra-sentential code-mixing on Devina Hermawan's YouTube channel. *Social Sciences, Humanities and Education Journal (SHE Journal)*, 3(2), 159–168.
- Situmorang, K. E., Pasaribu, T. K., & Sembiring, R. A. (2023). An analysis of code mixing and code switching on Najwa Shihab podcast “Susahnya Jadi Perempuan.” *ELT Worldwide: Journal of English Language Teaching*, 10(2), 356–366.
- Wardhaugh, R., & Fuller, J. M. (2015). *An introduction to sociolinguistics* (7th ed.). Wiley-Blackwell.
- Yuliana, N., Luziana, A. R., & Sarwendah, P. (2015). Code-mixing and code-switching of Indonesian celebrities: A comparative study. *Lingua Cultura*, 9(1), 47–54.