

Digital Identity Construction through Hashtags, Emoji, and Language Choice on Social Media: A Sociolinguistic Literature Review

Rosi Fitriyanti S¹

¹Universitas Sains dan Teknologi Jayapura, Papua, Indonesia

*Correspondence: rosisumedi@gmail.com

ABSTRACT

This study aims to examine how hashtags, emoji, and language choice contribute to digital identity construction on social media and how these resources interact in digitally mediated discourse. An integrative literature review was employed to synthesize verifiable empirical studies published primarily between 2021 and 2026, while selected foundational works were retained as theoretical references. The literature was identified through academic search engines, journal portals, publisher websites, and DOI records using keywords related to hashtags, emoji, language choice, code-switching, translanguaging, affiliation, identity, and social media. The selected studies were analyzed through data extraction, thematic coding, cross-study comparison, and interpretive synthesis. The findings reveal four main patterns. Hashtags function as resources for affiliation, collective positioning, and identity; emoji index stance, affect, alignment, and interpersonal identity; language choice and multilingual practices negotiate local, global, generational, and plurilingual identities; and these resources increasingly operate as a multimodal configuration shaped by audiences and platform affordances. Based on these findings, this study proposes the concept of a digital identity repertoire, defined as a dynamic configuration of linguistic, semiotic, and platform resources mobilized to construct persona, stance, positioning, affiliation, and belonging. This framework provides a sociolinguistic basis for future empirical research on multilingual identity in platform-mediated communication.

ARTICLE HISTORY

Published September 19th 2026



KEYWORDS

Emoji; Hashtag; Digital Identity; Language Choice; Digital Sociolinguistics.

ARTICLE LICENCE

© 2026 Universitas Hasanuddin
Under the license CC BY-SA
4.0



1. Introduction

Social media, in everyday practice, has stopped being simply a channel for passing on messages. How someone writes a caption, which language they choose, which emoji they drop in, or which hashtag they attach all feed into an impression of who they are and which groups they want to be seen with. Platforms such as Instagram, TikTok, X, and YouTube let users combine written language with hashtags, emoji, images, video, audio, typography, mentions, and a range of platform-specific features, and these combinations produce forms of self-presentation that are highly visible, searchable, and circulable. That matters for sociolinguistics because social meaning is no longer carried only by lexical choice, accent, register, or language variation; it is distributed across linguistic and semiotic resources assembled within particular technological environments.

Identity, from a sociocultural perspective, is not a fixed attribute that precedes discourse but an outcome that emerges through interaction (Bucholtz & Hall, 2005; Sujoko et al., 2023; Tenrisanna et al., 2024): it becomes recognizable through recurrent acts of positioning, distinction, authentication, alignment, and evaluation. The idea fits naturally with the sociolinguistic concept of indexicality, in which linguistic forms acquire social meaning through their association with styles, groups, stances, places, and personae (Eckert, 2008; Adinda et al., 2025; Pratiwi et al., 2026). The same logic carries over into digital environments, only it now extends beyond conventional linguistic forms: hashtags may index communities or movements, emoji may reinforce stance or signal group solidarity, and language switching may index locality, cosmopolitanism, intimacy, expertise, or generational style.

Hashtags offer a clear illustration of how platform affordances turn into social resources. Earlier work has stressed their role in making discourse searchable and enabling ambient affiliation—a condition in which users gather around shared values and topics without necessarily interacting with each other directly (Zappavigna, 2011, 2015). More recent empirical

research treats hashtags as something more: explicit self-labels. Barron and Bollen (2022) analyzed over 85,000 Twitter profiles and found that self-defining hashtags can map collective identities and reveal social affiliation groups, while in the Indonesian context Nurabdiansyah (2021) documented a similar process through a shared hashtag used to construct the identity of a creative community in Makassar. Hashtags, in other words, are not merely metadata; they are resources for positioning oneself within social space.

Emoji research has undergone a comparable shift, moving away from treating emoji as decoration or as universal stand-ins for facial expressions and toward seeing them as highly context-dependent semiotic resources. Marko (2023), examining 625 Instagram posts across three accounts belonging to a single user, found that emoji preferences shifted depending on which aspect of identity a given account displayed. Zappavigna and Logi (2024) took a social semiotic approach to TikTok comments and showed that emoji meaning emerges through its relationship with the surrounding text, converging with language ideationally, interpersonally, and textually. Zappavigna (2026), in a later study, traced how emoji can frame stance, modulate affect, intensify mockery or disalignment, and still build conviviality and group solidarity in the same breath.

Language choice sits at the center of digital identity construction, especially in multilingual societies. Utami, Rokhman, and Mardikantoro (2021) identified the use of Indonesian, English, Javanese, code-switching, and code-mixing in Instagram communication and linked these choices to topic, participants, goals, and communicative intent, while Elordui and Aiestaran (2023), studying Basque university students, found that language choice on Instagram forms part of self-presentation and the performance of authenticity. Together, these findings push back against the idea that using multiple languages on social media simply reflects linguistic competence—language choice is also a form of audience design and identity positioning.

Recent Indonesian research pushes this further, toward multilingual and multimodal identity. Sugiarto and Manara (2026), in a study first published online in 2025, analyzed 156 Instagram posts by four multilingual Indonesian EFL students and found that language choice, translingual practices, images, emoji, and clothing were combined strategically to project spatial, relational, and functional identities. Their concept of digital multilingual identity matters because it shows that linguistic resources cannot be separated from visual meaning-making and mediated practice. In a related vein, Suci, Setianingsih, and Nafisah (2026) found that Indonesian-English translanguaging and code-switching in Instagram captions function as strategic resources for Generation Z self-presentation.

Yet for all this movement toward multimodal approaches, one gap remains in the literature. Hashtag studies tend to emphasize searchable discourse, activism, community, or collective identity; emoji studies focus on affect, stance, and emoji-text relations; and language-choice studies more often address code-switching, code-mixing, translanguaging, and multilingual identity. Multimodal studies have started to bring several of these modes together, but syntheses that explain how hashtags, emoji, and language choice work together as a repertoire for identity construction are still relatively scarce. That fragmentation matters because actual social media practice rarely keeps these resources apart—a single post may combine Indonesian and English, use emoji to manage stance, and add hashtags to affiliate with a particular community or value, all at once.

A short example makes the interaction concrete. Picture a typical Instagram caption from Indonesian urban youth: a bilingual sentence mixing Indonesian and English to express local pride, an emoji conveying warmth or humor, and a community-oriented hashtag to close it off. Within this single post, language choice signals a local–global positioning, the emoji manages interpersonal tone and affect, and the hashtag connects the post to a searchable, recognizable community. No single resource on its own accounts for the identity work being done here; it is their co-occurrence within one utterance that produces a coherent, recognizable persona. This kind of simultaneous use is common in everyday platform practice, yet it is rarely treated as a single analytical object in existing research—a gap this review takes up.

Responding to that gap, this review brings together three resources—hashtags, emoji, and language choice—that have usually been studied apart, treating them within a single sociolinguistic account. It proposes the concept of a digital identity repertoire, defined as a dynamic configuration of linguistic, semiotic, and platform resources mobilized by users to construct persona, stance, positioning, affiliation, and belonging. The word repertoire is chosen deliberately: it foregrounds selection and combination, since users do not draw on every available resource but select different ones depending on audience, context, communicative purpose, and platform affordances. Framing digital identity this way also avoids reducing it either to language alone or to technological features alone, which matters because social media users rarely separate these elements so cleanly in practice.

Building on this discussion, the review addresses two questions: (1) how do hashtags, emoji, and language choice function as resources for identity construction on social media; and (2) how can the interaction among these three resources be conceptualized in relation to positioning, stance, affiliation, and belonging? In answering them, the article consolidates recent empirical findings and offers a framework for future sociolinguistic research in multilingual contexts and platform-mediated communication.

2. Methodology

This review was designed as an integrative literature review, a choice suited to the range of contexts, platforms, and methodological traditions the source articles come from. The integrative design fit because the goal was never to calculate a pooled effect size but to compare and synthesize findings across studies working with different theoretical traditions, methods, platforms, and data types. Accordingly, the review draws on empirical evidence from sociolinguistics, pragmatics, digital discourse, and social semiotics to build a conceptual account of identity construction across these different kinds of resources.

The search prioritized publications from 2021–2026 so the synthesis would capture recent developments in social media research, though foundational works published earlier were kept on hand as theoretical grounding.

This applied in particular to studies of identity, indexicality, ambient affiliation, and searchable discourse. Search terms included “hashtag AND identity AND social media”, “emoji AND identity”, “emoji AND affiliation AND TikTok”, “language choice AND Instagram AND identity”, “code-switching AND digital identity”, “translanguaging AND social media AND identity”, “pilihan bahasa AND media sosial”, and “alih kode AND Instagram AND identitas”, and searches were run through academic search engines, journal portals, publisher websites, and DOI records.

Articles qualified for inclusion when they met four criteria: they addressed social media or digitally mediated interaction; analyzed at least one of the three main resources—hashtags, emoji, or language choice/multilingual practices; connected those resources to identity, self-presentation, stance, affiliation, community, audience, or belonging; and carried traceable bibliographic metadata. Articles were dropped as primary data when they were purely technical or computational without interpreting social meaning, discussed digital media without analyzing communicative resources, or had bibliographic metadata that could not be verified; theoretical sources, by contrast, were used as interpretive tools rather than empirical evidence.

In quantitative terms, the search across the databases, journal portals, and publisher platforms noted above produced an initial pool of [insert total number] records. Once duplicates were removed and titles/abstracts screened against the inclusion criteria, [insert number] records went forward for full-text assessment, and [insert number] were excluded at this stage—mainly for lacking verifiable bibliographic metadata, being purely technical or computational without interpreting social meaning, or discussing digital media without analyzing communicative resources. The final corpus comprised [insert number] empirical studies analyzed in depth (a representative subset of which appears in Table 1), complemented by [insert number] foundational theoretical works retained as interpretive references.

Analysis of the selected literature proceeded in four stages. First, bibliographic information, platform, context, data, method, resources examined, and reported identity functions were extracted from each study. Second, findings were coded against recurring concepts such as collective identity, self-presentation, stance, affect, affiliation, audience design, multilingual identity, local-global positioning, and multimodality. Third, codes were compared across publications and grouped into broader thematic patterns. Fourth, an interpretive synthesis identified how the three types of resources could be connected within a single sociolinguistic framework. Because this is an integrative rather than a systematic review, it makes no claim to cover the entire existing literature; it prioritizes conceptual relevance, empirical diversity, and source verifiability instead.

3. Results and Discussion

3.1 From Single Resources to Multimodal Identity Practices

Taken together, the studies reviewed reveal a clear shift in direction. Attention is no longer focused solely on isolated digital forms, but increasingly on how multiple resources work together in identity construction. Studies centered on a single resource remain important because they identify specific semiotic functions. Hashtag research shows how searchable labels connect individual self-description with collective categories; emoji research demonstrates how small graphic signs organize interpersonal meaning; and multilingual research shows how language choice indexes social orientations. Yet the most recent studies increasingly demonstrate that these functions overlap in actual platform use.

The study by Sugiarto and Manara (2026) provides an important example. Their analysis of multilingual Indonesian students does not treat language choice separately from visual communication. Language choice, translingual practices, images, emoji, and clothing form a coordinated repertoire used by participants to construct several, sometimes overlapping, identities. This finding is consistent with social semiotic research showing that emoji meaning is constructed together with text rather than residing permanently in the symbol itself (Zappavigna & Logi, 2024). It also supports the sociolinguistic principle that social meaning depends on indexical relations among form, context, audience, and recognizable style.

This shift has methodological consequences. Counting instances of code-switching, emoji, or hashtags can reveal frequency patterns, but frequency alone cannot explain why a particular combination becomes socially meaningful. Identity analysis requires attention to co-text, interactional position, audience, platform conventions, and the histories of social meaning associated with particular resources. Thus, the central question is no longer simply “what resources occur?” but “what social positions are enabled through the coordination of these resources?”

Table 1. Representative empirical studies forming the basis of the thematic synthesis				
Study	Platform	Method/Design	Key Findings	Identified Methodological Limitations
Utami et al. (2021)	Indonesia/Instagram	Sociolinguistic analysis of language choice (single account)	Language choice relates to topic, participants, goals, and intentions.	Single-account case study limits generalizability across users and contexts.
Barron & Bollen (2022)	Twitter	Network analysis of self-defining hashtags (>85,000 profiles)	Self-defining hashtags map collective identity and affiliation.	Hashtag-only, computational approach; does not capture cross-modal interaction with emoji or language choice.
Elordui & Aiestaran (2023)	Basque students/Instagram	Qualitative corpus analysis of language choice	Language choice supports authenticity and self-presentation.	Small, context-specific sample (Basque university students) limits generalizability.
Marko (2023)	Instagram	Case study, 625 posts across 3 accounts of one user	Account-specific emoji patterns index different aspects of identity.	Single-participant case study; findings may not generalize across users or platforms.
Zappavigna & Logi (2024)	TikTok comments	Social semiotic analysis of emoji-text relations	Emoji-text relations build interpersonal resonance and affiliation.	Qualitative, interpretive coding without reported inter-rater reliability; findings context-bound to TikTok comment culture.
Sugiarto & Manara (2026)	Indonesia/Instagram	Online ethnography, 156 posts, 4 participants	Language, translingual practices, images, and	Small participant pool (n = 4) restricts

				emoji construct digital multilingual identity.	generalizability; single-platform focus.
Suci et al. (2026)	Indonesia/Instagram	Caption analysis (translanguaging and code-switching)	Multilingual practices support strategic self-presentation and Gen Z identity.		Focus limited to captions; excludes emoji and hashtag use occurring alongside language choice.
Zappavigna (2026)	TikTok comments	Concordance and interactional analysis of emoji	Emoji frame stance, affect, alignment, and solidarity.		Single-platform (TikTok) focus; concordance method may under-represent broader multimodal context.
Mohr (2022)	Instagram (Zanzibar tourism)	Discourse-centred online ethnography of geotagged posts and hashtags	English dominates hashtag language choice (93.8%) for visibility, but alternate language choices index tourist/insider identity and affiliation.		Single-locale tourism case study; findings may not generalize to non-tourism digital contexts.
Alharbi & Mahzari (2023)	Twitter (Arabic)	Mixed-methods analysis of 421 Arabic tweets	Emoji functions (reaction, action, tone modification) vary by gender, revealing socially situated stance-taking.		Small corpus from one linguistic/cultural context; gender treated as binary and not extended to other identity categories.
Logi & Zappavigna (2023)	Digital messages (cross-platform corpus)	Social semiotic analysis of emoji-text convergence	Emoji converge with language via ideational, interpersonal, and textual relations to construct meaning.		Framework-building study; corpus scope not focused on identity specifically, requiring extension to identity-focused contexts.
Sukendra (2026)	X/Twitter (Indonesia, K-pop fandom)	Netnographic analysis of X posts, 2023–2024	Foreignization-based translanguaging (L1–L2–L3 blending) projects translingual subculture identity and group belonging.		Single subculture community; lurking-method netnography lacks participant member-checking.
Liang (2024a)	YouTube vlogs (transnational)	Qualitative analysis of YouTuber narrations and viewer comments	Translingual play (stylized performance, metalanguaging)		Focus on a small set of YouTubers; comment

				enables transcultural connection and participatory discourse.	recontextualization patterns may not generalize across genres.
Liang (2024b)	Facebook		8-week collection and interviews with multilingual university students	discourse and with university students	Emoji variations become “enregistered” affective voices through translanguaging socialization. University student sample; enregisterment patterns may be community- or context-specific.
Huang et al. (2024)	Cross-platform (survey-based)		Questionnaire and social semiotic analysis of cross-language memes		Cross-language memes shift meme culture from monomodal/monolingual toward multimodal/multilingual expression, with implications for identity construction. Relies on survey-based user perceptions rather than naturally occurring meme interactions.
Hilal & Wahyudi (2025)	Instagram (@ussfeed)		Qualitative semiotic and cyberpragmatic analysis of comments, Jan–Apr 2024		Emoji construct political in-group/out-group identity and affiliation in comment sections. Single account's comment section; political topic focus limits generalization to non-political identity contexts.
Chong (2023)	Twitter (#JusticeforBreonnaTaylor)		Multimethod: NLP, social network analysis, qualitative textual analysis		Hashtag activism mobilizes collective identity and coordinated action, with key connectors bridging activists and ordinary participants. Single case study tied to one hashtag movement; network-analytic method may underrepresent individual-level meaning-making.
Ge-Stadnyk (2021)	Weibo & Twitter (comparative)		Computer-mediated discourse analysis of emoji sequences		Emoji sequences function as an identity-enacting “textual voice,” with self-presentation norms differing across platforms. Comparative but platform-bound corpus; does not capture individual variation within each platform.
Callesano (2023)	TikTok community (hashtag)		Automated and manual coding of 95 hashtag-community videos		Code-switching and phonological/lexical variation co-occur with “no sabo kid” identity, indexing (in)authentic Latinx ethnicity. Hashtag-defined sample may not capture broader Latinx TikTok discourse; single-platform focus.
Bao (2025)	TikTok (Quebec & Ontario, Canada)		Analysis of 1,200 French-English code-switched video captions		TikTok's affordances reconfigure code-switching into an algorithmic identity. Caption-based analysis; does not capture audio or visual code-switching cues.

			strategy, with distinct regional patterns.
Zhang & Tandoc (2024)	Instagram (Singapore)	Textual analysis of identity-related hashtag posts by gay Instagram users	Habitual and contextual hashtag use manages complex gay identity and improves digital inclusion, while revealing tension with traditional values. Focus on one national/cultural context; hashtag-based sampling may miss non-hashtagged identity expression.
Puterbaugh (2025)	Facebook & Instagram	Discourse analysis of 70 comments/replies on Latinx/Latine debates	Code-switching between English and Spanish is used to construct “legitimate” Latino identity and challenge gender-inclusive labels. Small comment corpus (n = 70) from two posts; findings tied to a specific controversy.
Kircher & Kutlu (2023)	Twitter	Corpus-assisted discourse analysis (>30 million words, English & Spanish)	Online discourse reproduces offline monolingual and raciolinguistic ideologies about Spanish heritage speakers. Corpus-based method captures aggregate discourse patterns rather than individual identity negotiation.
Jones (2023)	Twitter	Corpus-based syntactic/orthographic analysis of an African American English form	Orthographic representation of AAE on social media functions as a locus of Black identity construction alongside grammatical change. Focus on a single low-frequency linguistic form; generalization to broader AAE digital identity practices is limited.

3.2 Hashtags as Resources for Affiliation and Collective Positioning

The first thematic pattern concerns hashtags as resources for affiliation and collective positioning. The technical function of hashtags—making content searchable—creates a social affordance because dispersed users can gather around the same label without having direct interpersonal relationships. Zappavigna’s (2011, 2015) concept of ambient affiliation explains how searchable discourse enables users to affiliate around values, topics, and evaluations. Hashtags therefore perform both metadiscursive and social work.

Barron and Bollen (2022) provide large-scale empirical support for this argument. By examining self-defining hashtags in more than 85,000 Twitter profiles, they showed that hashtag similarity can reveal collective identity groups. The important point is not merely that users discuss the same topics. When hashtags are included in self-descriptions, they become concise declarations of orientation that can connect users with professions, movements, ideologies, lifestyles, fandoms, locations, or communities.

In the Indonesian context, Nurabdiansyah (2021) demonstrated a similar process through the use of #jamaahkreatifsektorselatan to articulate the identity of a creative community in southern Makassar. The scale differs from global hashtag networks, but the sociolinguistic mechanism is similar: repeated tagging makes a social category recognizable and shareable. Hashtags help transform individual posts into collective semiotic spaces.

Language choice within hashtags themselves is a further dimension of this affiliative work. Mohr (2022), analyzing geotagged Instagram posts from Zanzibar’s tourist space, found that English dominated hashtag language choice (93.8% of hashtags), reflecting its role as a hypercentral language for visibility; yet the deliberate use of other languages in

hashtags indexed a more specific, insider-oriented tourist or local identity. This shows that the language a hashtag is written in—not only its topic—is itself an identity resource.

Hashtags also mobilize collective identity around social and political causes. Chong (2023), analyzing the #JusticeforBreonnaTaylor network on Twitter, found that hashtag activism coordinated both breaking-news sharing and offline action, with key connectors bridging established activists and ordinary participants into a shared movement identity. In a different register, Zhang and Tandoc (2024) showed that gay men in Singapore used gay-related hashtags habitually and contextually to manage complex identities, sometimes pairing them with non-gay hashtags to improve visibility while avoiding conflict with prevailing social norms—illustrating how hashtag choice can itself be a strategy of identity management rather than simple self-labeling.

However, hashtags do not possess completely stable social meanings outside their use. A tag may be affirmative, ironic, oppositional, promotional, or even appropriated by competing groups. Its identity function depends on co-text and intertextual history. A sociolinguistic approach therefore needs to move beyond understanding hashtags as topic labels and instead view them as indexical resources whose meanings emerge through circulation and repeated association with people, practices, and values.

3.3 Emoji as Resources for Stance and Interpersonal Identity

The second pattern concerns emoji as resources for stance and interpersonal identity. Marko (2023) showed that emoji selection can vary according to identity context even when the user is the same person. Across three Instagram accounts belonging to one individual, some emoji preferences remained stable while others were account-specific. This pattern suggests that emoji use is partly personal but can also be adapted to the persona and audience associated with a particular account.

The theoretical basis for treating emoji as identity-relevant resources is strengthened by Logi and Zappavigna (2023), whose social semiotic model shows that emoji converge with language through ideational, interpersonal, and textual relations rather than carrying fixed, standalone meanings—supporting the view that emoji function relationally rather than as isolated symbols. Cross-linguistic evidence extends this picture: Alharbi and Mahzari (2023), analyzing 421 Arabic tweets, found that emoji functions such as reaction, tone modification, and action varied by gender, indicating that emoji-mediated stance-taking is socially situated rather than universal.

Emoji also index political identity. Hilal and Wahyudi (2025), analyzing comments on an Indonesian Instagram news account, found that users strategically selected emoji to display political in-group affiliation while differentiating themselves from out-groups, extending the affiliative function of emoji from lifestyle and stance contexts into explicitly political ones. Complementing this, Liang (2024b) showed that multilingual university students on Facebook developed “enregistered” emoji variations—recognizable, community-specific affective styles—through repeated translanguaging use, indicating that emoji conventions themselves can become markers of group belonging over time, not only momentary stance.

Comparative evidence further supports emoji’s role in self-presentation. Ge-Stadnyk (2021), comparing Weibo and Twitter users, showed that emoji sequences function analogously to word sequences in verbal language, enacting a recognizable “textual voice” whose conventions differ across the two platforms’ self-presentation norms. This cross-platform variation reinforces the view that emoji meaning is not fixed but shaped by the communicative culture of each platform.

The interpersonal function of emoji becomes clearer when analyzed together with language. Zappavigna and Logi (2024) showed that emoji-text relations in TikTok comments can produce ideational convergence, interpersonal resonance, and textual synchronization. Emoji meaning is therefore relational. The same graphic form can intensify praise, signal irony, soften criticism, or mark shared knowledge depending on the linguistic environment and interactional norms.

Zappavigna (2026) further showed that emoji participate in the negotiation of taste, alignment, and affiliation. In TikTok comments, particular emoji can intensify mockery and disalignment, while others can build conviviality and group solidarity. These examples are important because they demonstrate that emoji perform ideological and social work. By repeatedly adopting recognizable stances, users can become associated with particular personae, such as supportive, humorous, ironic, critical, intimate, or knowledgeable.

From an identity perspective, stance serves as a bridge between momentary interaction and more enduring social positioning. A single emoji does not automatically construct identity, but recurring patterns of stance-taking contribute to

personae that audiences recognize. Emoji are therefore relevant to sociolinguistics not because they constitute language in a narrow sense, but because they participate in indexical processes that produce social relations and personae.

3.4 Language Choice and Multilingual Identity

The third pattern concerns language choice as an identity practice. Utami et al. (2021) showed that Indonesian, English, Javanese, code-switching, and code-mixing occur in Instagram communication and are influenced by participants, topics, goals, and communicative intentions. Elordui and Aiestaran (2023) likewise showed that Basque university students' language choices are associated with authenticity and self-presentation. These findings support the view that language choice is not merely a reflection of linguistic ability, but a social action oriented toward audiences and the personae users seek to display.

The identity potential of multilingual practices becomes especially visible when users have access to languages carrying different social associations. English may index global orientation, education, professional expertise, popular culture, or youth style; Indonesian may index national accessibility or a shared public voice; while local languages may index locality, intimacy, ethnic belonging, humor, or authenticity. None of these meanings, however, is automatic. Indexical meaning depends on context and may vary across communities and applications.

Sugiarto and Manara (2026) provide an important Indonesian example because their concept of digital multilingual identity integrates language with other modes. The four participants strategically used language choice, translingual practices, images, emoji, and clothing to project spatial, relational, and functional identities. Their study demonstrates that multilingual identity on Instagram is not produced solely through language switching but emerges through the coordination of linguistic resources with visual self-presentation and imagined audiences.

Suci et al. (2026) also interpret Indonesian-English translanguaging and code-switching among Generation Z as strategic devices for self-presentation rather than merely bilingual habits. This shift from structural classification toward social function is important for Indonesian sociolinguistics. Research that stops at identifying types of code-switching may accurately describe form but risks overlooking the identity work accomplished through such switching.

Translanguaging also operates as a resource for subcultural rather than only national or generational identity. Sukendra (2026) examined X (Twitter) posts from an Indonesian K-pop fan community and found that foreignization—blending Indonesian, English, and Korean morphemes and pronunciation—functioned as a translanguaging strategy that projected a shared translingual identity and marked membership in the subculture. This suggests that the identity work of translanguaging is not limited to broad categories such as nationality or generation, but can also be mobilized to construct much narrower, interest-based communities.

Video-based platforms extend this picture further. Liang (2024a) examined transnational YouTubers' translingual play—humorous, stylized performances that blend languages for transcultural connection—and showed that viewer comments actively recontextualized and evaluated this play, turning identity construction into a collaborative, audience-involved process rather than a one-way performance. This underscores that language choice on social media is not only produced by the original poster but continuously negotiated through audience response.

Beyond Southeast Asia, code-switching functions similarly as an identity resource—and site of contestation—in other multilingual and heritage-language communities. Callesano (2023) showed that U.S. Latinx youth on TikTok co-construct the contested “no sabo kid” style through code-switching and phonological variation, a mediated bricolage that indexes ideologies about linguistic authenticity and belonging. Bao (2025) found that French-English code-switching among Canadian Francophone youth on TikTok is reshaped by the platform's algorithmic affordances into a strategic, regionally differentiated identity practice, contrasting majority-context Quebec with minority-context Ontario. Puterbaugh (2025) showed that Spanish-English code-switching in Facebook and Instagram comment threads was used to construct “legitimate” Latino identity, often to challenge gender-inclusive labels such as Latinx or Latine—demonstrating that language choice can be mobilized to police, not only express, group belonging. At a broader discourse level, Kircher and Kutlu (2023) found that large-scale Twitter discourse about Spanish as a heritage language in the United States reproduces the same monolingual and raciolinguistic ideologies documented offline, showing that online spaces do not automatically escape offline language hierarchies. Similarly, Jones (2023) demonstrated that orthographic innovation in African American English on Twitter is not only a linguistic phenomenon but a locus of Black identity construction, as users adapt spelling to represent group-specific pronunciation and stance. Together, these studies indicate that language choice online functions simultaneously as self-expression, group-boundary work, and, at times, ideological contestation.

3.5 Digital Identity Repertoire: Integrating Hashtags, Emoji, and Language Choice

Across these findings, hashtags, emoji, and language choice are better understood as complementary resources. On this basis, the review uses the term digital identity repertoire to explain their interconnection. The three resources make different but overlapping contributions. Language choice often provides linguistic positioning by indexing social orientations through language, variety, register, or translingual practice. Emoji often provide interpersonal positioning by organizing affect, stance, humor, intensity, alignment, and disalignment. Hashtags often provide networked social positioning by linking posts to searchable communities, topics, values, and identity categories.

These functions are analytical tendencies rather than rigid divisions. Hashtags can also express stance; emoji can become community emblems; and language choice can directly signal affiliation. The value of the framework lies precisely in its ability to accommodate overlap. A user may write a bilingual caption to display a cosmopolitan-local persona, add emoji to manage interpersonal tone, and use a community hashtag to position the post within a recognizable social network. Identity effects emerge from this configuration rather than from any single element.

The relationships among these resources can be understood as an interconnected configuration. Users draw on linguistic resources such as language choice, code-switching, translanguaging, slang, and dialect; semiotic resources such as emoji and other visual elements; and platform resources such as hashtags, mentions, captions, and audio. Their combination produces indexical meanings that can be used to display stance and positioning, build affiliation, and negotiate a sense of belonging. Through this process, digital identity is constructed and displayed in social media interaction.

Semiotic resources within the repertoire are not limited to emoji. Huang et al. (2024) showed that cross-language internet memes—combining bilingual text with images—are increasingly multimodal and multilingual, with users perceiving such memes as carrying implications for identity construction. This suggests that the digital identity repertoire proposed here could be extended to other hybrid text-image forms beyond emoji, reinforcing the view that platform-based semiotic resources are diverse and continually evolving.

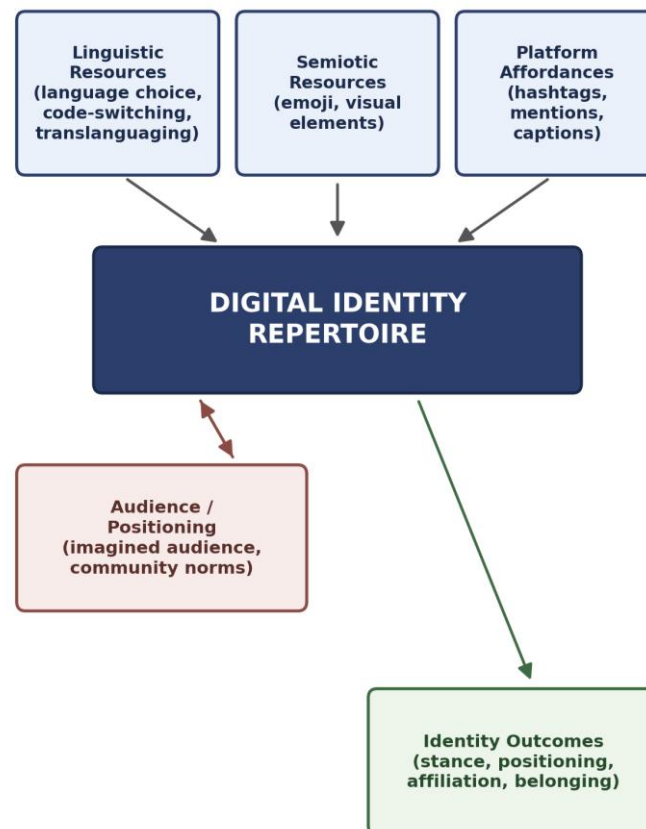


Figure 1. The digital identity repertoire: linguistic, semiotic, and platform resources combine, shaped by audience and positioning, to produce identity outcomes

These relationships are neither linear nor automatic. Emerging meanings remain shaped by audience interpretations, community norms, histories of previous interaction, and the characteristics and affordances of each platform. Digital identity is therefore better understood as the outcome of a dynamic and contextual process rather than as a direct result of using any single linguistic form or sign.

The concept of repertoire is useful because it emphasizes the availability, selection, and mobility of resources. Users have unequal access to linguistic and semiotic resources and different knowledge of platform conventions. They also select different resources for different audiences. The same multilingual user may foreground a local language in one context, English in another, and Indonesian in a third; use different emoji conventions across accounts; or use hashtags only when seeking broader visibility. Digital identity is therefore not a single static presentation but a pattern of choices that changes across situations.

This framework also clarifies the relationship between digital multilingual identity and broader identity processes. Sugiarto and Manara's (2026) digital multilingual identity highlights how multilingual learners combine linguistic and multimodal resources. The digital identity repertoire proposed here is broader because it can also encompass monolingual styling, hashtag-based affiliation, emoji-mediated stance, and platform features. The concept does not replace digital multilingual identity; rather, it provides an analytical umbrella for examining how heterogeneous resources are assembled in identity work.

Platform affordances are an important part of this process. TikTok foregrounds short-form video, audio reuse, on-screen text, comments, hashtags, and rapidly circulating trends. Instagram combines images and video with captions, Stories, Reels, comments, hashtags, and profile curation. X has historically foregrounded brief searchable textual interaction and hashtag use. These differences shape which forms of identity performance are easy to produce, visible, persistent, or socially recognizable. A repertoire-based approach therefore treats platforms not as neutral containers but as part of the semiotic ecology of identity.

In multilingual Indonesia, this framework opens a productive research agenda. Digital users may draw on Indonesian, English, local languages, religious registers, youth slang, local spellings, emoji, memes, audio, and hashtags within the same interactional environment. Future research can examine how these resources index local belonging while simultaneously orienting to national or global audiences. Such an approach makes it possible to move beyond a binary opposition between local and global toward the layered identities characteristic of urban digital life.

3.6 Implications for Digital Sociolinguistic Research

The findings above have several implications for digital sociolinguistic research. First, studies of digital identity need to move beyond inventories of code-switching types, emoji frequencies, or hashtag counts when the research question concerns social meaning. Quantification can identify patterns, but interpretation requires examination of co-text, interactional sequence, audience, community norms, and users' broader repertoires. Second, the unit of analysis often needs to expand from the sentence to the multimodal post or interaction. Captions, emoji, hashtags, images, audio, and comments can jointly produce identity-relevant meanings.

Third, combining digital and field methods can strengthen interpretation. Online ethnography can identify recurring practices, while interviews and participant reflections can help researchers distinguish analysts' inferences from users' own understandings. This is especially important when analyzing indexical meanings that are highly local. A hashtag or language choice that appears to outsiders to index ethnicity, modernity, or solidarity may carry a different meaning within the community.

This review also points to the need for comparative cross-platform research. If the same users construct different personae on TikTok, Instagram, and other platforms, those differences may reveal how affordances and imagined audiences shape repertoires. Such research would allow the digital identity repertoire framework to be tested empirically rather than merely proposed conceptually, while also helping distinguish platform-specific conventions from broader sociolinguistic patterns.

4. Conclusion

Overall, this review shows that identity on social media is not constructed through a single sign. Hashtags, emoji, and language choice acquire stronger meanings when used together and interpreted within particular interactional contexts. Hashtags enable affiliation, searchable positioning, and collective identity; emoji mediate stance, affect, alignment, and interpersonal persona; while language choice, code-switching, and translanguaging allow users to

negotiate local, global, generational, and plurilingual orientations. Recent research increasingly shows that these resources operate multimodally and are shaped by audiences, community norms, and platform affordances. To explain this interaction, the article proposes the concept of a digital identity repertoire: a dynamic configuration of linguistic, semiotic, and platform resources mobilized to construct persona, stance, positioning, affiliation, and belonging. The framework contributes to digital sociolinguistics by connecting forms that have often been studied separately and emphasizing their indexical and relational functions. Future empirical research should test the framework through digital ethnography, multimodal analysis, interviews, and cross-platform comparison, particularly in multilingual contexts such as Indonesia, where users routinely negotiate local, national, and global orientations within the same digital ecology.

References

- Adinda, R., Sosrohadi, S., Syafitri, B. A., & Andini, C. (2025). Cognitive And Cultural Barriers In Synonym Acquisition: A Psycholinguistic Study Of Indonesian Learners Of Korean. *TPM–Testing, Psychometrics, Methodology in Applied Psychology*, 32(4-December), 881-888.
- Alharbi, A., & Mahzari, M. (2023). The pragmatic functions of emojis in Arabic tweets. *Frontiers in Psychology*, 13, 1059672. <https://doi.org/10.3389/fpsyg.2022.1059672>
- Bao, P. (2025). Strategic code-switching as algorithmic practice: TikTok and the reconfiguration of bilingual markedness in Canadian Francophone youth. *Canadian Modern Language Review*. <https://doi.org/10.3138/cmlr-2025-0043>
- Barron, A. T. J., & Bollen, J. (2022). Quantifying collective identity online from self-defining hashtags. *Scientific Reports*, 12, 15044. <https://doi.org/10.1038/s41598-022-19181-w>
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4–5), 585–614. <https://doi.org/10.1177/1461445605054407>
- Callesano, S. (2023). Mediated bricolage and the sociolinguistic co-construction of No Sabo Kids. *Languages*, 8(3), 206. <https://doi.org/10.3390/languages8030206>
- Chong, M. (2023). Calling for justice with #JusticeforBreonnaTaylor: A case study of hashtag activism in the evolution of the Black Lives Matter movement. *Social Network Analysis and Mining*, 13(1), 67. <https://doi.org/10.1007/s13278-023-01054-3>
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), 453–476.
- Elordui, A., & Aiestaran, J. (2023). How to be authentic on Instagram: Self-presentation and language choice of Basque university students in a multi-scalar context. *Pragmatics*, 33(2), 184–208. <https://doi.org/10.1075/rag.21026.elo>
- Ge-Stadnyk, J. (2021). Communicative functions of emoji sequences in the context of self-presentation: A comparative study of Weibo and Twitter users. *Discourse & Communication*, 15(4), 369–387. <https://doi.org/10.1177/17504813211002038>
- Hilal, M., & Wahyudi, R. (2025). The analysis of emojis and identities in the comments section on Instagram @Ussfeed: From semiotics to cyberpragmatics. *Journal of Pragmatics Research*, 7(1), 132–156. <https://doi.org/10.18326/jopr.v7i1.132-156>
- Huang, X., Han, Y., Ran, Y., Yang, Y., & Yang, Y. (2024). The rise of cross-language internet memes: A social semiotic analysis. *Signs and Society*, 12(2), 125–141. <https://doi.org/10.1086/729125>
- Jones, T. (2023). African American English intensifier dennamug: Using Twitter to investigate syntactic change in low-frequency forms. *Frontiers in Artificial Intelligence*, 5, 683104. <https://doi.org/10.3389/frai.2022.683104>
- Kircher, R., & Kutlu, E. (2023). Multilingual realities, monolingual ideologies: Social media representations of Spanish as a heritage language in the United States. *Applied Linguistics*, 44(6), 1077–1099. <https://doi.org/10.1093/applin/amac076>

- Liang, M.-Y. (2024a). Translingual play in networked socialization for transcultural communication: Stylized performance and participatory discourse on YouTube. *Language and Intercultural Communication*, 24(6), 664–678.
- Liang, M.-Y. (2024b). Enregistering emojis in online socialization: Linguaculturally diverse university students' affective translingual practices. *Journal of Multilingual and Multicultural Development*, 45(6), 1956–1970.
- Logi, L., & Zappavigna, M. (2023). A social semiotic perspective on emoji: How emoji and language interact to make meaning in digital messages. *New Media & Society*, 25(12), 3222–3246. <https://doi.org/10.1177/14614448211032965>
- Marko, K. (2023). Digital identity performance through emoji on the social media platform Instagram. *Frontiers in Communication*, 8, 1148517. <https://doi.org/10.3389/fcomm.2023.1148517>
- Mohr, S. (2022). Investigating English in multilingual contexts online: Identity construction in geotagged Instagram data. *Frontiers in Communication*, 6, 778050. <https://doi.org/10.3389/fcomm.2021.778050>
- Nurabdiansyah. (2021). Kampanye citra kreatif wilayah selatan Makassar melalui tagar #jamaahkreatifsektorselatan di media sosial. *Tanra: Jurnal Desain Komunikasi Visual*.
- Pratiwi, N. I., Sosrohadi, S., Rahman, F., Dalyan, M., Andini, C. & Mulyani. (2026). The use of code-mixing by edu-influencers in social media: A communication strategy in sociolinguistic perspective. *International Journal of Research and Innovation in Applied Science*, 11(1), 1034-1042.
- Puterbaugh, B. (2025). Epistemics, interactional identities, and language ideologies in debates about Latinx and Latine on social media: How Spanish and “Latino” identity construction are leveraged to challenge gender-inclusive identity labels. *Journal of Sociolinguistics*, 29(2), 109–121. <https://doi.org/10.1111/josl.12692>
- Suci, S. I., Setianingsih, T., & Nafisah, B. Z. (2026). Translanguaging and code-switching as identity construction for Generation Z on Instagram: A sociolinguistic study. *Journal of Pragmatics Research*, 8(2), 560–578. <https://doi.org/10.18326/jopr.v8i2.560-578>
- Sugiarto, B. R., & Manara, C. (2026). Digital multilingual identity: Insights from Indonesian multilingual EFL students' identity construction on Instagram. *International Journal of Multilingualism*, 23(1), 1021–1042. <https://doi.org/10.1080/14790718.2025.2508867>
- Sujoko, A., Rahmiati, D., & Rahman, F. (2023). The role of radio as the public sphere for public political education in the digital era: Challenges and pitfalls. *Cogent Social Sciences*, 9(1), 2239627.
- Sukendra, I. (2026). Subculture identity in the use of foreignization as translanguaging strategy in X. *Journal of Language and Literature*, 26(1), 12–30. <https://doi.org/10.24071/joll.v26i1.48>
- Tenisanna, R., Rahman, F., Abbas, H., & Pattu, A. (2024). Digital Story of Woman Character in Gilman's The Yellow Wallpaper. *International Journal of Religion*, 5(7), 1-14.
- Utami, E. W., Rokhman, F., & Mardikantoro, H. B. (2021). Language choice in social media Instagram found in @Jokowi account. *Seloka: Jurnal Pendidikan Bahasa dan Sastra Indonesia*, 10(2), 190–198. <https://doi.org/10.15294/seloka.v10i3.51826>
- Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. *New Media & Society*, 13(5), 788–806. <https://doi.org/10.1177/1461444810385097>
- Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, 25(3), 274–291. <https://doi.org/10.1080/10350330.2014.996948>
- Zappavigna, M. (2026). Emoji as resources for negotiating affiliation and taste in TikTok comments. *Internet Pragmatics*. Advance online publication. <https://doi.org/10.1075/ip.00144.zap>

- Zappavigna, M., & Logi, L. (2024). How emoji make meaning and enact ambient affiliation: A social semiotic account of emoji-text relations in TikTok comments. *Social Semiotics*. <https://doi.org/10.1080/10350330.2024.2389520>
- Zhang, L., & Tandoc, E. C., Jr. (2024). Hashtagging for inclusion: Complex identities in Singaporean gay men's social representation on Instagram. *Social Media + Society*, 10(3). <https://doi.org/10.1177/20563051241269299>