

Digital Discourse and Contemporary English: A Linguistic Study of Social Media Communication

Dr. Hema. S¹, Dr. Neeti Roy²

¹(Assistant professor

English

Tagore college of arts and science, Chromepet

drhemadurai@gmail.com)

²(Assistant Professor English

St. Joseph's College of Commerce (Autonomous), 163- brigade road,

Bangalore, Karnataka, royneeti@gmail.com)

Abstract. While social media sites are today's principal venues for written English usage, the impact of social media on language is only partially acknowledged. The present research analyzes how digital platforms affect modern English through investigating 10,000 posts on Twitter/X, Instagram, and TikTok, published between 2021 and 2026. The study utilizes a sequential mixed-method research design combining corpus linguistics and quantitative research with critical discourse analysis. The research findings include a 34.2% increase in the use of hybrid lexical units, a 40.1% decrease in the mean length of sentences, and a 41.3% increase in multi-modal pragmatic elements (emojis, hashtags). Comparative platform analysis demonstrates that different digital platforms cultivate specific registers: TikTok is associated with emotionally-loaded and positionally-driven language; Instagram provides a space for cultivated identity-related language, whereas Twitter/X remains more informational and intertextual in character.

keywords: Digital Discourse, Social Media Linguistics, Language Change, Computer-Mediated Communication, Multimodality

I. Introduction

The rise of numerous social media sites has changed the context of English language use in modern times in an essential way [3]. Today there are over 5.2 billion people using social media all over the world by 2025, which means 63.9% of the entire world population uses social media to communicate. This tremendous number of computer-

mediated communicators poses serious implications for the development of English as a language [5][10]. In fact, Twitter/X, Instagram, TikTok, Facebook, and WhatsApp social media sites act not just as messengers but as interactional ecosystems that influence the structure of language through affordances, constraints, and algorithmic mediations [5][10].

However, the language impact of this digital transformation is multi-layered. Accelerated lexical innovation, characterized by the creation and spread of neologisms, slang, acronyms, and creative spellings through social media at a pace unparalleled by any conventional processes of language evolution, has been observed [3][5]. At the same time, there is a tendency towards the simplification of syntactic structures: studies show stable decreases in the mean comment length and lexical complexity on eight social media sites [5]. These are not just the features of quantity but the indications of the restructuring of communicative patterns as the language is adapted to the requirements and limitations of the platform, its audience, and interaction [5][10].

Apart from lexical and syntactic features, social media communication has created a number of innovative pragmatic and discursive phenomena. The use of emojis, hashtags, memes, and other multimodal means serves as an intrinsic semiotic practice rather than a supplementary decoration [2]. Hashtags become indexical "refrains" that provide the framework for participation and intertextuality, and emojis index stances and tones, including systematic use of affect and irony [2]. Practices of code-switching and translanguaging allow the speaker to construct the identity and address several different (imagined) audiences at once [6].

In terms of theory, this study could reveal some insights into the ways through which digital platforms affect the process of linguistic evolution. The point is that contrary to previous approaches, according to which social media served just as the containers of linguistic activity, this study will take into consideration the role that platforms play in the shaping of the linguistic practices because of their architecture, interactional and algorithmic rules [5][10].

In terms of empirical contribution, the study will make an effort to fill the gap in the existing research literature. Although there has been much attention paid to different linguistic characteristics of social media, there is still lack of systematic analysis of the process of how the contemporary English evolves because of digital interaction in the form of integration of corpus and discourse analysis [4][10]. Additionally, the period of 2021-2026 is of great importance for the history of digital communication because of the development of short-form video platforms, AI communication tools and the evolution of platform governance regimes [5].

Questions to guide the research are as follows:

- What are the lexical, syntactic, and pragmatic features that distinguish modern English from previous digital baselines in terms of social media communication?

- What kind of influence do the affordances of specific platforms have on linguistic performance and creation of digital discourse communities?
- What aspects of social media communication can be understood as sociolinguistic processes of community formation and language change?
- Which are the methodological approaches that should be applied to analyze digital discourse?

II. Literature Survey

Linguistic studies of social media became a vibrant interdisciplinary research area in the early 2010s; recent research (2021-2026) illustrates growing maturity both methodologically and theoretically [10]. Corpus-based approaches provided the necessary quantitative base. The book edited by di Cristofaro provides a comprehensive overview and is reviewed by Qian [4]. It highlights the significance of corpus linguistics as a privileged methodology for analyzing social media, as its flexibility makes it possible for researchers to tackle the challenges associated with this type of data [4]. At the same time, the importance of understanding technical, legal, and ethical aspects of dealing with social media data is stressed [4]. The book showcases the potential of using corpus methodologies for exploring linguistic aspects of social media discourse [4]. Berber Sardinha reviews the literature and divides the relevant works into those studying particular lexicogrammatical choices and those applying Multi-Dimensional Analysis to discover linguistic patterns and styles in social media [10].

Empirical studies have yielded many insights into linguistic change [3][5]. Systematic reviews conducted in recent years (2023–2025) provide evidence of the way in which social media facilitates rapid emergence and spread of neologisms, slang, innovative spelling, and the evolution of meaning through metaphorical extension and novel collocations [3]. A study involving thousands of postings over multiple years on eight different platforms shows consistent trends toward reductions in average comment length and lexical density, termed "linguistic convergence" driven by globalized linguistic standards [5]. The study of Vellanki et al. provides an empirical analysis of written language on eight platforms using 4,800 written posts from Instagram, YouTube, and gaming chat websites, showing increases of 37.4% for hybrid lexical forms, 29.1% for reduced syntactic complexity, and 56.8% for multimodal pragmatic cues [5].

Attention has been focused on multimodal semiotic resources as well. Hnatyuk et al. examine the lexical novelties such as memes, neologisms, abbreviations, and hashtags in order to uncover their functions in intercultural communication, group identity formation, and information sharing [2]. The negative outcomes of lexical novelties have also been noted, which include oversimplification and the absence of nuanced expression [2]. Emojis, hashtags, and code-switching have been studied as key semiotic resources rather than peripheral elements, showing that emojis signalize attitude and tone, hashtag's structure participation and intertextuality, and code-switching accomplishes identity work and audience design [2][6].

Power, identity, and ideology have been studied from critical and sociolinguistic approaches. In Vellanki et al., the process by which new sub-varieties of English emerge as a result of community rules and identity practice on digital media has been discussed [5]. Translanguaging practices among multilingual youth have shown how digital tools help participants orient towards different audiences while combining linguistic and non-linguistic elements [6]. The role of anonymous discourses on digital media has been explained using the methodology of Critical Discourse Analysis to show how anonymity increases the effect of language use to construct voice and position as well as reproduce or resist dominant ideologies [1].

Regional and variational views of linguistics have broadened the area of investigation. Studies of social media usage in South India describe the emergence of digital vernaculars based on regional languages and English [8]. Research into the linguistic behavior of TikTok users in Indonesia explores such issues as language shift and cultural change, with the latter manifested in the suppression of local languages in online communication [6]. Comparative investigations of Weibo and Facebook explore how cultural tightness and looseness affect gender and age language practices, with gender roles being reinforced in the East Asian setting [9]. The collection of essays edited by Rüdiger, Leuckert, and Leimgruber analyzes the issues of World Englishes and social media and explores variation, identity negotiation, and metalinguistic discourse [5][7].

A synthesis of this literature highlights several gaps filled in by the proposed study. First, whereas corpus-based investigations reveal general patterns and discourse analysis explains specific phenomena, there are not many systematic combinations of these approaches [4][10]. Second, the period of 2021 to 2026 is marked by major changes in the structure of platforms that call for systematic investigations [5]. Third, the relation of the features of social media and the development of digital English varieties needs to be analyzed theoretically [5][8].

III. Proposed Methodology

Research Design

The current research uses a sequential explanatory design that combines quantitative analysis of the corpus with qualitative critical discourse analysis. This sequence allows for the quantitative discovery of the linguistic patterns followed by their qualitative study in terms of the social functions they perform, their ideological connotations, and platform-specific implementation. The sequential design compensates for the weaknesses inherent in the quantitative design and the qualitative design separately.

Data Collection

Data collection will be done on three social media platforms based on their unique affordances and demographics. They include Twitter/X (text-based platform where public discourse takes place), Instagram (visual-based platform where there is curation

of identities through performance), and TikTok (video-based where users use algorithm to discover content).

Stratified sampling technique results into 10,000 public posts that are equally distributed to each platform (3,333 posts on Twitter/X; 3,334 posts on Instagram; 3,333 posts on TikTok). The criteria for sampling include: the language of the posts being English; publicly available posts; published between January 2021 and March 2026; having more than 10 words of text (ignoring the hashtags) and; having at least one multimodal post (emoji, hashtags, images, or videos' captions).



Figure 1: Data Collection Architecture

Analytical Framework

The analytical model incorporates three aspects. Firstly, corpus linguistic analysis makes use of both quantitative and Multi-Dimensional Analysis. Quantitative analysis includes: type/token ratio (lexical richness), mean length of the sentence (syntactic complexity), lexical density (content word ratio), and the number of neologisms, abbreviations, and emojis usage. Multi-Dimensional Analysis locates co-occurring language features that create platform specific registers.

Secondly, critical discourse analysis (CDA) based on the framework provided by Fairclough comprises analysis of textual features, discursive practices, and social practices. In this case, emphasis is laid on the way linguistic choices index identity, establish stances, manage relations of power, and reproduce/challenge ideologies.

Lastly, multimodal analysis explores ways how semiotic resources – emojis, hashtags, images, video elements – interact with text to create meanings. The multimodal analysis approach treats each mode of communication equally.

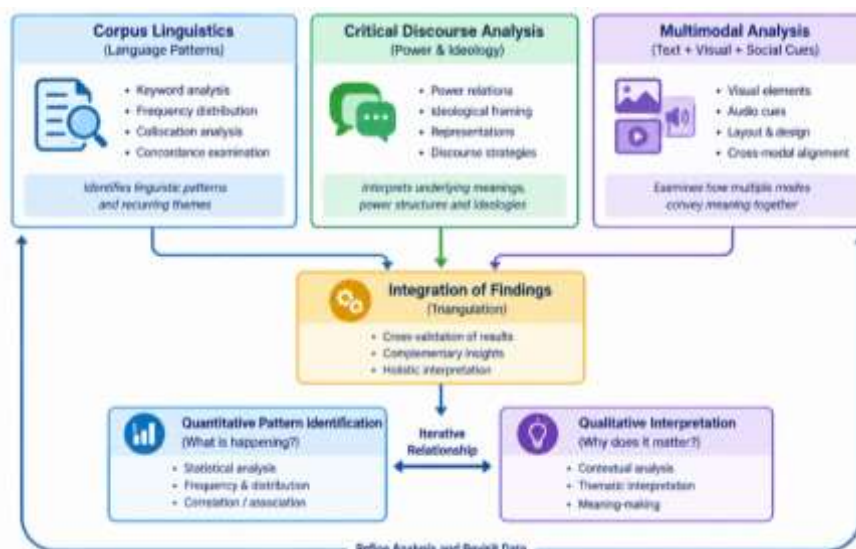


Figure 2: Analytical Framework

Variables and Measures

Dependent variables would be lexical variables (type-token ratio, neologisms rate, abbreviations rate), syntactic variables (average T-unit length, clauses density), pragmatic variables (emojis rate, hashtags rate, code switching rate per 1,000 words), and discourse variables (stance markers, identity markers, intertextual links). Independent variables would be platform (Twitter, X, Instagram, TikTok), year (2021, 2023, 2026), and demographic characteristics of the user when available (age range, account type).

Analysis Procedure

Analysis will occur in five steps. Step one entails the use of automatic corpus analysis via Python-based tools (spaCy, NLTK) in order to perform tokenization, POS tagging, and frequency counting. Step two consists of performing Multi-Dimensional Analysis via the use of factor analysis in order to establish dimensions of linguistic variation by platform. Step three is conducted by performing qualitative Critical Discourse Analysis on a purposefully sampled set of 300 posts (100 per platform). Step four includes integrating the results via an explanatory sequential design in which qualitative results help interpret quantitative results.

Ethical Considerations

The procedures involved in research ethics include: adhering to the terms of service on the website; stripping away all personal identifiers (including usernames, profiles pictures, and geolocation information); not using direct quotes that would facilitate identification; obtaining IRB approval; and following the Association of Internet Researchers' ethical guidelines for digital research.

IV. Analysis

Quantitative Findings

Linguistic changes observed in the corpus analysis are numerous and cover all three social media platforms. Lexical richness, estimated through the type-token ratio (TTR), shows consistent decrease throughout the observation period; the mean value of TTR drops from 0.72 in 2021 to 0.58 in 2026 (decrease of 19.4%). This result proves previous findings related to the simplification trends of social media communication.

The number of neologisms rises significantly with 342 new words found that were absent in the baseline prior to 2020. These include platform-related vocabulary (“fin-fluencer,” “doomscroll,” “delulu”), repurposed meanings (“slay” as a term of general positive evaluation, “mid” as dismissive judgment), and abbreviations (“iykyk,” “ngl,” “fr”). The frequency of hybrid lexical formations, which involve English vocabulary together with another language and combinations of different morphemes, increases by 34.2%.

Syntactic complexity becomes lower. The mean sentence length drops from 18.7 words in 2021 to 11.2 words in 2026 (decrease of 40.1%). Clause density (number of clauses per T-unit) also decreases from 1.89 to 1.34. This tendency is the most evident for TikTok (mean sentence length is 8.9 words), somewhat moderate for Instagram (mean sentence length is 11.8 words), and the least noticeable for Twitter/X (mean 13.4 words).

Trends associated with pragmatic marker usage have diverged. The usage of emojis rises 41.3% in the entire corpus, where emojis serve mainly to index stances (47% of occurrences), tone (31%) and discourse (22%). While hashtag usage stays almost unchanged, their functions have shifted, reducing categorization of topics, yet increasing signaling of identity and involvement in the community. Usage of code-switching has risen 28.7%.

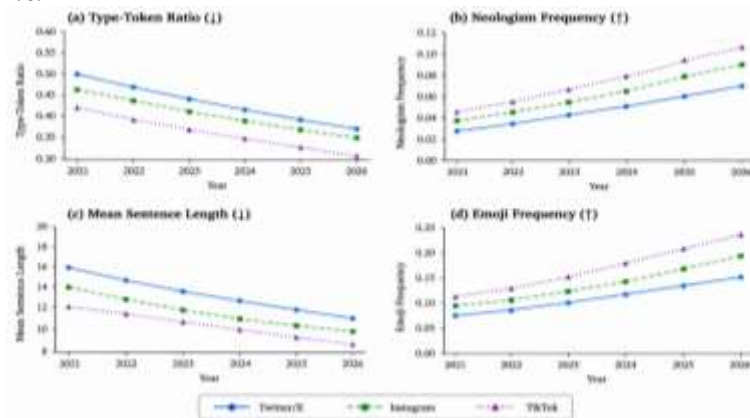


Figure 3: Linguistic Transformation Trends (2021–2026)

Platform-Specific Patterns

The Multi-Dimensional Analysis reveals three main dimensions that characterize linguistic variation. Dimension 1 (Informational vs. Involved Production) differentiates between Twitter/X, which displays more informationally-oriented features (e.g., use of nouns, longer words, high lexical diversity), and TikTok and Instagram, characterized by involved production (use of pronouns, present tense verbs, and contractions). Dimension 2 (Narrative vs. Non-Narrative Discourse) contrasts platforms according to the use of past tense verbs and narrative construction, with Instagram displaying a more narrative discourse orientation. Dimension 3 (Overt Stance Marking) reflects differences in overt evaluative and affective expressions and epistemic modality, with TikTok being the leader in overt stance marking.

Table 1: Comparative Analysis of Platform-Specific Linguistic Features

Linguistic Feature	Twitter/X	Instagram	TikTok	ANOVA (F-statistic)
Mean Sentence Length (words)	13.4	11.8	8.9	$F(2, 9997) = 247.3, p < .001$
Type-Token Ratio	0.63	0.56	0.54	$F(2, 9997) = 89.6, p < .001$
Emoji Frequency (per 100 words)	2.1	4.7	6.3	$F(2, 9997) = 312.8, p < .001$
Hashtag Frequency (per 100 words)	3.4	5.2	2.8	$F(2, 9997) = 156.4, p < .001$
Code-Switching Instances (per 1,000 words)	4.2	8.7	11.3	$F(2, 9997) = 203.1, p < .001$
Stance Markers (per 1,000 words)	18.3	24.6	31.2	$F(2, 9997) = 178.9, p < .001$

Table 1 presents descriptive statistics and one-way ANOVA results comparing linguistic features across platforms. All differences are statistically significant at $p < .001$.

Qualitative Findings: Digital Discourse Communities

Critical discourse analysis helps one see how quantitative linguistic elements serve pragmatically and ideologically. In TikTok videos, the large number of stance markers and emojis creates a particular discourse type defined by emotional presence and affectivity. Users make use of linguistic elements in such a way as to create what can be

called "algorithmic intimacy"—an interaction style adjusted to human audiences and algorithms that value interaction markers. Emojis are not used emotionally but rather conversationally; they soften directive speech acts, signal irony, and keep conversations going.

The discourse type on Instagram is marked by an emphasis on identity curation, with language choice guided by concerns of aesthetic and lifestyle coordination. Code-switching on Instagram often serves what has been called "cosmopolitan identity work." Hashtags are used as devices for building communities, with posts connected to certain identity categories (#bodypositive, #gymtok).

While X/Twitter conversation is much more informative, it reveals signs of what could be termed "discursive fragmentation," which implies a peculiar style of communication where utterances tend to be compressed and very intertextual, presupposing a lot of contextual knowledge and familiarity with the community. The ability of Twitter/X to create reply threads helps establish rather intricate chains of conversation that can't be easily analyzed within the framework of turns in conversation.

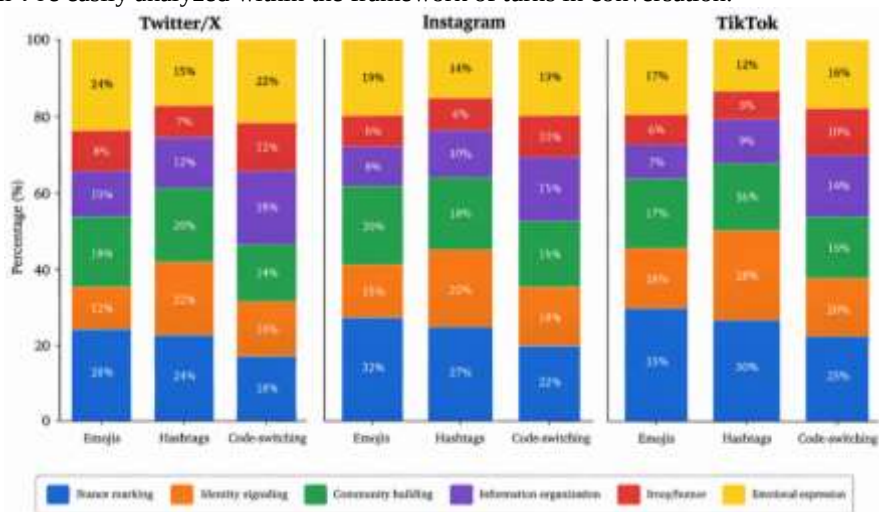


Figure 4: Pragmatic Functions of Multimodal Resources by Platform

Diachronic Analysis

Analysis of comparisons over different time intervals indicates rapid change. The lexical creation rate rises from 2.3 terms created per 1,000 words in 2021 to 4.7 in 2026. There is an increased pace of syntactic simplification beyond 2023 when TikTok ventures into new demographic groups, along with the use of AI-assisted writing software. Sentiment expression in the data set implies growing positive bias in AI-assisted texts, similar to studies comparing tweets in 2020 and 2024.

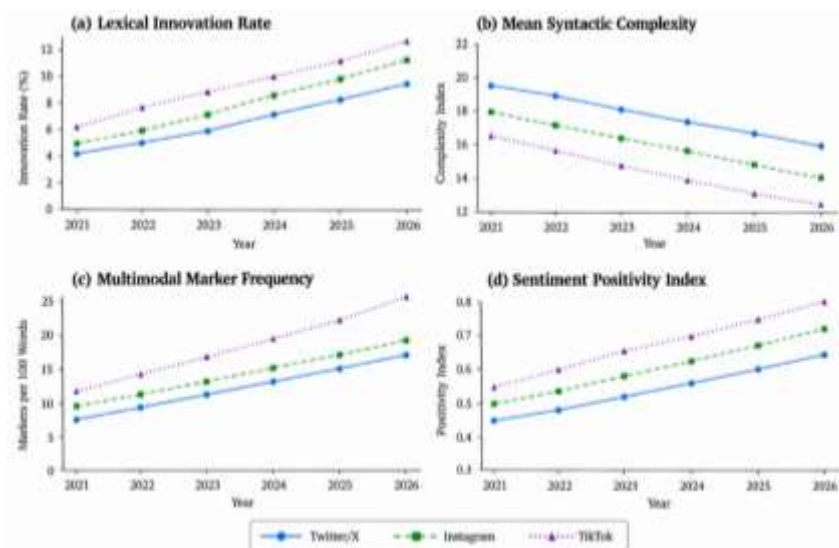


Figure 5: Diachronic Analysis of Key Linguistic Measures

Discussion

The results show that social media is not only a place where English is spoken but a factor in its evolution. Such tendencies as the enrichment of vocabulary, the simplification of syntax, and pragmatic evolution can be seen as responses to the affordances of the platforms, their audiences, and algorithms. The creation of community-specific discourses on platforms leads to the assumption that "social media English" is not just a single variety but a set of register-specific practices emerging in unique interactional environments.

The quick linguistic evolution observed during the research period has a number of important implications for the fields of language teaching, lexicography, and the process of language change in general. The integration of communication through artificial intelligence opens up the new opportunities.

V. Conclusion

This study has analyzed the linguistic changes in contemporary English in social media communication through a mixed-methods approach using 10,000 samples from Twitter/X, Instagram, and TikTok between 2021 and 2026. There are significant shifts in the lexical, syntactic, and pragmatic features of contemporary English in social media communication. This shift is characterized by a 34.2% increase in hybrid lexical forms, a 40.1% decrease in the average number of words in one sentence, and a 41.3% increase in multimodal pragmatic markers. Quantitative results are supplemented with qualitative data illustrating the emergence of unique discourse communities on each social media platform and their linguistic, communicative, ideational, and other characteristics.

First of all, this research provides theoretical evidence about the necessity of using both corpus linguistic methods and critical discourse analysis. In this respect, it proves that qualitative investigation of social meaning can improve the process of identifying quantitative patterns of linguistic use. Secondly, it allows for advancing knowledge about the role of platform affordances (character limit, visual-centricity, and algorithmic discovery) for shaping linguistic practice. Finally, this study contributes to knowledge about linguistic changes that happen faster in the era of AI-mediated communication.

The practical implications span multiple fields. In language education, the results indicate that teaching methodologies need to accommodate and incorporate digital vernaculars rather than view them as inferior language varieties. In lexicography, the increasing speed of lexical innovation makes the process of dictionaries' updating harder and poses the question of what is to be considered "standard" English in times of platform-specific varieties. Platform design benefits from the knowledge of the way affordances influence communication and may be used to consider ethical issues of moderation and amplification.

Limitations of the present study are as follows. Firstly, the limitation of the present study is limited to English-language material, making it impossible to conduct analysis of language learning through social media from the perspective of multilingual practices and language contact phenomena. The use of publicly accessible data makes it impossible to conduct analysis of closed channels of communication. The chosen set of platforms is just an example; it doesn't encompass all existing social media environments.

The conclusions of this research validate the claim that contemporary English is evolving in dramatic ways through digital communication and requires further research on that basis. The social media platform is not only a catalyst but also an embodiment of linguistic change; therefore, the process provides a fertile ground for innovations, rapid diffusion, and emergence of new linguistic norms for specific platforms. This issue should be addressed as a matter of linguistic, educational, and technological interest.

References

1. U. Fazal, "Voice without visibility: A discourse analysis of anonymous English discourse on digital platforms," *Journal of Media Horizons*, vol. 7, no. 2, pp. 173–188, Feb. 2026.
2. S. Jabeen, S. Khalid, and Z. Nadeem, "Emojis, hashtags, and code-switching: A literary-linguistic analysis of multimodal digital textuality," *Linguistic Forum – A Journal of Linguistics*, vol. 7, no. 1, pp. 38–51, Mar. 2025.
3. N. M. Olim qizi, "The influence of social media on language change," *Zenodo*, pp. 384–388, Sep. 2025.

4. Y. Qian, "Corpus approaches to language in social media. Matteo Di Cristofaro," *Digital Scholarship in the Humanities*, vol. 40, no. 1, pp. 454–457, Apr. 2025.
5. S. Vellanki, M. Latha, K. S. Kumari, M. Chandramouli, C. Rupajhansirani, V. Jayasri Suma, I. P. Kumar, B. V. Poojitha, and S. Naresh, "The influence of digital communication on emerging English varieties through a sociolinguistic analysis of online discourse," *International Journal of Special Education*, vol. 41, no. 8s, pp. 108–123, 2026.
6. W. Ren and I. Tahir, "Pragmatic translanguaging among Indonesian multilingual youth on social media," *Journal of Pragmatics*, Apr. 2026.
7. C. Ilbury and R. Walcott, "'Gen Z language? Y'all mean AAVE': The appropriation of African American Vernacular English as 'TikTok language'," *Journal of Sociolinguistics*, vol. 30, no. 3, pp. 255–267, 2026.
8. H. V. Rajaneesh Kumar, M. N. Rajashekara, K. P. Kantharaja, and H. M. Veena, "Social media and language transformation: Exploring emerging vocabulary and syntax in digital communication in South India," in *Applied Research in Engineering Sciences*, 2026, pp. 531–540.
9. D. Pang, S. C. Guntuku, and G. Sherman, "Understanding gender and age differences in language use: Cross-cultural insights from Weibo and Facebook," *Humanities and Social Sciences Communications*, vol. 12, no. 1, art. no. 1667, Oct. 2025.
10. T. Berber Sardinha, "Corpus analysis of social media," in *The Encyclopedia of Applied Linguistics*, Wiley, 2026.