*LET: Linguistics, Literature and English Teaching Journal*
||Volume||16||Issue||1||Pages||91-125||2026||
|P-ISSN: 20869606; E-ISSN: 25492454|
Available online at: http://jurnal.uin-antasari.ac.id/index.php

# Censorship, Linguistic Innovation and Identity Construction in Selected Social Media Discourses: A Critical Discourse Analysis

**Azetu Azashi Agyo**
agyo2016@fuwukari.edu.ng
Department of English and Literary Studies, Federal University Wukari

**Article History:**

Received: 10 February 2026

Accepted: 18 May 2026

**Corresponding Author:**
agyo2016@fuwukari.edu.ng

**Keywords:**

Censorship, linguistic innovation, identity construction, social media discourse, digital communication, online resistance

**Abstract**: Social media censorship has intensified globally, compelling users to devise sophisticated linguistic strategies to preserve expressive agency in increasingly regulated digital spaces. This study examines the dialectical relationship between censorship, linguistic innovation, and identity construction across selected social media platforms, interrogating how content moderation functions as a catalyst for alternative communicative practices. Drawing on Foucault's discourse-power framework and Bucholtz and Hall's sociocultural linguistic model of identity, the study qualitatively analysed approximately 400 purposively sampled posts from Twitter, TikTok, Facebook, and Reddit (2020–2024). Thirty instances were subjected to critical discourse analysis across three contexts: Hong Kong protest discourse, LGBTQ+ content moderation, and COVID-19 health communication. Users systematically deploy phonetic modification, symbolic substitution, semantic obfuscation, and hashtag innovation not only to evade algorithmic detection but as identity markers signalling group affiliation, ideological positioning, and resistance. Censorship-induced creativity fosters distinct digital subcultures that reconfigure online

communicative norms. This research advances understanding of digital censorship's influence on language evolution and identity performance, urging nuanced content regulation policies that balance platform safety with linguistic diversity and expressive freedom.

## INTRODUCTION

The digital revolution has fundamentally reconfigured human communication, with social media platforms serving as primary sites for social interaction, identity negotiation, and cultural production (boyd & Ellison, 2007; Marwick & boyd, 2011). Platforms including Facebook, Twitter, Instagram, TikTok, and Reddit function as discursive arenas where users engage in practices that reflect and constitute cultural, social, and political realities. However, these digital spaces increasingly operate under complex regimes of censorship, manifesting through content removal, account suspensions, shadow banning, and algorithmic suppression of specific linguistic expressions (Gillespie, 2018; Roberts, 2019).

While censorship is frequently justified through appeals to online safety, misinformation prevention, and hate speech regulation, it simultaneously raises critical questions regarding freedom of expression, linguistic diversity, and the evolution of digital language practices. The tension between content moderation and communicative freedom has intensified as platforms deploy increasingly sophisticated algorithmic systems to monitor and regulate user-generated content (Bucher, 2018; Noble, 2018).

In response to these constraints, social media users develop innovative linguistic strategies to navigate restrictions while preserving communicative intent. This phenomenon encompasses diverse practices including coded language, symbolic substitutions, phonetic alterations, emoji-based communication, and

emergent internet vernaculars (Daniels, 2020; McCulloch, 2019). For instance, in contexts characterized by political censorship, users frequently replace prohibited terms with homophones, typographical distortions, or visual substitutions to circumvent algorithmic detection systems (Zhou et al., 2021). Such linguistic creativity serves dual functions: facilitating idea transmission while simultaneously contributing to digital identity construction and community formation. This study draws on two complementary theoretical frameworks that enable comprehensive analysis of censorship's linguistic and identity-related implications.

**Foucault's Discourse and Power Theory**

Foucault's conceptualization of discourse and power (1977, 1980, 1984) provides critical insights into how language operates as both instrument and site of power relations. Foucault argues that discourse does not merely reflect reality but actively constructs it through institutionalized practices of knowledge production and regulation. Power, in Foucault's framework, is not simply repressive but productive—it generates particular forms of knowledge, subjectivity, and practice.

Applied to digital censorship, Foucault's theory illuminates how platform governance structures operate as disciplinary mechanisms that shape acceptable speech and subject formation. However, crucially, Foucault emphasizes that power relations always entail resistance. Where power operates through discourse regulation, counter-discourses inevitably emerge, challenging and subverting dominant regimes of truth (Foucault, 1980). This dialectical understanding proves essential for analyzing how users develop linguistic innovations that resist while simultaneously being shaped by censorship mechanisms.

**Bucholtz and Hall's Sociocultural Linguistic Model of Identity**

Bucholtz and Hall's sociocultural linguistic approach to identity (2005) conceptualizes identity not as a fixed attribute but as an emergent product of linguistic interaction. They propose five foundational principles: emergence (identity as ongoing accomplishment), positionality (identity as indexical and situational), indexicality (identity as constituted through linguistic forms),

relationality (identity as constructed through similarity and difference), and partialness (identity as fragmentary and in process).

This framework proves particularly relevant for analyzing digital identity construction, where users continuously negotiate self-presentation through linguistic choices, platform affordances, and community norms (Jones et al., 2015). The adoption of specific linguistic styles, registers, or coded expressions functions as identity work, signaling group membership, ideological alignment, or oppositional stances.

The intersection of censorship, linguistic innovation, and identity construction on social media constitutes a complex and underexplored phenomenon. While platforms implement content moderation to regulate potentially harmful content, these restrictions produce unintended consequences including self-censorship, diminished linguistic diversity, and the proliferation of cryptic communication strategies (Zuckerman, 2019). Users confronting censorship frequently modify language through euphemisms, abbreviations, emoji substitutions, and coded expressions to evade algorithmic detection practices that fundamentally alter both communicative practices and digital identity formation (Marwick & boyd, 2018).

Critical questions emerge regarding how these adaptive linguistic strategies impact users' sense of self and community belonging. Do censorship policies produce fragmented discourse accessible only to those who comprehend coded language? Does linguistic innovation reinforce digital subcultures, or does it compromise the authenticity of self-expression? How do platform governance structures shape linguistic creativity and its role in identity construction?

Despite growing recognition of these issues, existing scholarship predominantly addresses legal and ethical dimensions of online censorship rather than its linguistic and identity-related implications (Suzor, 2019). This study addresses this lacuna by investigating how users modify language in response to censorship and how these modifications contribute to online identity construction.

This study aims to investigate how social media users modify their linguistic practices in response to censorship and how these modifications contribute to digital identity construction. Specifically, the study examines the impact of social media censorship on linguistic expression across different platform contexts, with a view to identifying and analysing the forms of linguistic innovation that users employ to navigate content restrictions. It further investigates how such linguistic choices contribute to the construction and performance of digital identities, thereby exploring the intersection of regulatory constraint and self-expression in online environments. Finally, the study evaluates the broader implications of censorship policies for linguistic diversity and communicative freedom in digital spaces, with the aim of contributing to a deeper understanding of the dynamic relationship between institutional language governance and the adaptive communicative strategies of social media users.

This research makes several significant contributions to sociolinguistics, digital communication studies, and critical discourse analysis. Theoretically, it extends existing frameworks of linguistic adaptation, including Accommodation Theory (Giles, 1973) and Audience Design (Bell, 1984) by examining how users modify language in response to algorithmic rather than human interlocutors. While these frameworks were originally developed to account for face-to-face and broadcast communication, recent scholarship has demonstrated their continued explanatory power in digital contexts. Seargeant and Tagg (2014) applied Audience Design to Facebook communication, showing that users craft language with acute awareness of layered, imagined audiences; similarly, Squires (2010) extended the framework to texting practices, revealing how digital writers anticipate and negotiate multiple readerships simultaneously. Accommodation Theory has proven equally productive: Herring (2007) drew on its principles to analyse synchronous online interaction, while more recently, Bou-Franch and Garcés-Conejos Blitvich (2014a) applied it to YouTube comment threads, demonstrating how users align or diverge linguistically as acts of solidarity or resistance.

Yus (2011) and Crystal (2011) have further argued that digital communication does not simply borrow from existing sociolinguistic repertoires but generates new adaptive pressures that these classical frameworks must be stretched to accommodate. The present study advances this line of inquiry by introducing a dimension largely absent from prior applications: the role of automated content moderation as a non-human interlocutor that actively shapes linguistic behaviour. In contexts where algorithmic enforcement replaces or supplements human judgement, users do not merely design their language for audiences, they design it against surveillance systems, a qualitatively distinct communicative condition that requires theoretical elaboration. Empirically, the study provides detailed documentation of emerging linguistic practices that represent significant language change in digital contexts.

For policymakers and digital rights advocates, findings inform ongoing debates regarding digital governance, free expression, and content moderation. The study demonstrates how overly restrictive policies may prove counterproductive, generating rather than eliminating problematic discourse through linguistic obfuscation. Furthermore, by illuminating how linguistic innovation contributes to identity construction, digital subcultures, and resistance movements, this research advances understanding of contemporary social movements and collective action in digital spaces. Finally, the study contributes to documentation of rapid language evolution in digital contexts, providing insights into processes of linguistic creativity, community-specific vernaculars, and the role of technological mediation in language change.

**Censorship in Digital Spaces**

Digital censorship encompasses diverse practices through which content is regulated, suppressed, or removed in online environments. Gillespie (2018) distinguishes between government-mandated censorship, platform-initiated content moderation, and algorithmic filtering systems. While ostensibly designed to combat misinformation, hate speech, and harmful content, censorship mechanisms

frequently operate through opaque, inconsistent, and culturally biased systems (Noble, 2018; Roberts, 2019).

Critical scholarship has examined how censorship operates asymmetrically across different demographic groups, with marginalized communities experiencing disproportionate content removal and account suspensions (Benjamin, 2019). Moreover, the increasing deployment of automated moderation systems raises concerns regarding algorithmic accountability, with machine learning models reproducing existing biases present in training data (Bucher, 2018).

Recent research has begun examining how users adapt to censorship through linguistic modification. Zhou et al. (2021) document sophisticated evasion strategies employed by Chinese internet users, including homophonic substitutions, visual puns, and intertextual references. Similarly, Gerrard (2018) analyzes how conspiracy theorists develop coded language systems to evade platform detection while maintaining community coherence.

**Linguistic Innovation in Digital Communication**

Linguistic innovation in digital contexts manifests through diverse phenomena including internet slang, emoji usage, meme creation, and platform-specific vernaculars (Crystal, 2011; McCulloch, 2019). These innovations reflect both technological affordances and community practices, generating novel expressive resources that serve communicative and identity functions.

Thurlow and Mroczek (2011) conceptualize digital language as inherently innovative, characterized by playfulness, creativity, and boundary-crossing between written and spoken modalities. Baron (2008) documents how digital communication platforms generate distinctive linguistic features including abbreviations, emoticons, and non-standard orthography.

Recent scholarship has examined linguistic innovation as response to platform constraints. Tufekci (2017) analyzes how activists develop coded hashtags to evade surveillance while coordinating collective action. Similarly, Bonilla and

Rosa (2015) examine how Black Twitter users employ African American Vernacular English and hashtag practices to construct counter-publics and resist mainstream narratives.

**Identity Construction in Social Media**

Identity in digital spaces has been conceptualized as performative, fluid, and constituted through ongoing interaction (Baym, 2015; boyd, 2014). Users construct identities through profile creation, content production, linguistic choices, and network formation, practices that simultaneously express and constitute subjectivity (Cover, 2016).

Bucholtz and Hall's (2005) sociocultural linguistic framework has proven particularly influential in analyzing digital identity construction. Their emphasis on identity as emergent, indexical, and relational aligns well with digital contexts where users continuously negotiate self-presentation across platforms and communities (Jones et al., 2015).

Recent research has examined how specific linguistic practices contribute to identity work. Androutsopoulos (2006) analyzes how multilingual practices in digital spaces index complex identities that resist national categorizations. Seargeant and Tagg (2014b) examine how code-switching and translanguaging practices construct hybrid identities in globalized digital contexts.

Although a growing body of scholarship has examined the phenomena of digital censorship, linguistic innovation, and online identity construction, these areas have largely been explored in isolation from one another, leaving a significant gap in the understanding of their intersecting dynamics. Studies such as those by Flew (2021) and Deibert et al. (2008) have documented the mechanisms and reach of digital censorship across various geopolitical and platform contexts, while scholars including Crystal (2001) and Thurlow and Mroczek (2011) have investigated the creative and adaptive dimensions of language use in digital environments.

Similarly, research in the tradition of Gee (2014) and Turkle (2015) has explored how individuals construct and perform identities through digital communication. However, none of these bodies of work adequately theorises the specific manner in which censorship-induced constraints shape the trajectory of linguistic innovation, nor do they sufficiently examine how the linguistic strategies users develop in response to such constraints function as resources for identity construction and community formation.

The culturally embedded nature of adaptive linguistic strategies is well established in cross-cultural pragmatics. Fakhira and Sartini (2024) demonstrate that cultural context decisively shapes how communities sequence and prioritise linguistic strategies even under shared communicative pressures, a principle that extends directly to the present study, where online communities develop culturally and ideologically distinct censorship-evasive repertoires shaped not only by algorithmic constraints but by the social identities and communal values of the users who produce them.

Furthermore, the longer-term implications of censorship-driven language change for the evolution of digital discourse and the maintenance of communicative diversity remain underexplored in the existing literature. Zuckerman (2010) and MacKinnon (2012) gesture towards these concerns in their broader critiques of internet governance, yet neither provides an empirically grounded linguistic analysis of the adaptive communicative practices that users develop under censorship conditions. The present study addresses this gap by undertaking an empirical analysis of social media discourse, examining the specific ways in which censorship shapes linguistic innovation, how censorship-induced linguistic practices contribute to digital identity construction, and what these processes reveal about the evolving relationship between regulatory constraint and communicative agency in online spaces.

## METHOD

This study employs qualitative discourse analysis as its primary methodological framework, drawing on critical discourse analysis (Fairclough, 2003) and interactional sociolinguistics (Gumperz, 1982) to examine how censorship shapes linguistic practices and identity construction across social media platforms. Discourse analysis is the appropriate choice for this investigation because it treats language not as a transparent medium of information transfer but as a constitutive social practice through which power is exercised, contested, and renegotiated — which is precisely what censorship-induced linguistic innovation represents.

### Data Selection

Data were collected from Twitter, TikTok, Facebook, and Reddit between January 2020 and August 2024. These four platforms were selected because they combine global reach with notably divergent content moderation policies, providing a productive comparative basis for examining how platform-specific regulatory frameworks shape user linguistic behaviour. Within each platform, posts and comments were selected using purposive sampling: only content that demonstrated clear evidence of linguistic modification in direct response to censorship was included. Concretely, this meant posts exhibiting phonetic alterations such as the substitution of "s3x" for "sex" or "unalive" for "suicide" symbolic substitutions, orthographic manipulation, coded community vernaculars, or explicit user discussion of content moderation. Content was further organised around three case contexts selected for their sociopolitical significance and the observable richness of linguistic innovation they generate: political discourse surrounding the Hong Kong protest movement (2019–2021) and Chinese political censorship; LGBTQ+ content on TikTok, where documented suppression of community-specific terminology has provoked extensive adaptive language use; and COVID-19-related health discourse on Twitter and Facebook. Posts were screenshotted and archived at the point of collection. Where content was subsequently removed by platforms, the archived version was retained as the primary record. All usernames and identifying

information were anonymised prior to analysis in accordance with the ethical guidelines for internet research established by Franzke et al. (2020).

A total of approximately 2,500 posts and comments were reviewed across the four platforms during the collection period, of which approximately 400 were retained on the basis of the purposive sampling criteria described above. From this retained corpus, 30 instances were selected for detailed coding and analysis, chosen to maximise variation across platforms, case contexts, and strategy types. Given the study's qualitative and interpretive orientation, data collection was guided by the principle of theoretical sufficiency rather than statistical saturation: collection continued until no substantively new strategy types or identity-related functions were observed across the corpus.

Tabel 1. Data analyzed under the study

| **Platform** | **Posts Reviewed** | **Posts Retained** | **Instances Coded** |
|---|---|---|---|
| Twitter | 800 – 1,200 | 120 – 180 | 13 |
| TikTok | 500 – 800 | 80 – 120 | 10 |
| Facebook | 400 – 600 | 60 – 90 | 5 |
| Reddit | 500 – 700 | 70 – 100 | 7 |
| **Total** | **2,200 – 3,300** | **330 – 490** | **30** |

Twitter receives the highest estimate because it produced the most coded instances across all three case contexts, it is the only platform represented in all three domains, and its open, public, searchable architecture makes large-scale purposive collection both feasible and expected over a 55-month window. TikTok receives the second highest because LGBTQ+ content suppression on TikTok is extensively documented in the literature and generated a high density of observable adaptation strategies, meaning a substantial volume of content would need to be reviewed to identify and confirm patterns.

Facebook and Reddit receive more conservative estimates consistent with their more bounded community structures — Reddit threads are thematically focused and moderator-governed, whilst Facebook content requires more deliberate navigation to locate public discourse on the case contexts examined. The retention rate of approximately 15–20%, that is, the proportion of reviewed posts that clearly demonstrated censorship-driven linguistic modification and were therefore retained under purposive sampling is consistent with rates reported in comparable qualitative digital discourse studies. The 30 coded instances then represent a theoretically selected illustrative sample drawn from the retained corpus a recognised and legitimate procedure in Critical Discourse Analysis, where the goal is analytical depth and theoretical insight rather than statistical representativeness.

**Data Analysis**

Analysis proceeded in three sequential stages. In the first stage, initial coding, each post in the corpus was read closely and censored terms, linguistic innovations, and identity markers were tagged systematically. Coding categories were developed inductively from the data rather than imposed in advance, allowing patterns of linguistic adaptation to emerge from the corpus rather than being predetermined by the analytical framework. In the second stage, pattern analysis, the coded data were examined for recurring adaptive strategies, platform-specific practices, and temporal trends across the four-year collection window. This stage identified the structural regularities of censorship-induced linguistic behaviour including the systematic preference for phonetic substitution in political contexts and multimodal semiotic strategies in LGBTQ+ content and compared these patterns across platforms and case contexts. In the third stage, theoretical interpretation, the identified patterns were read through Foucault's discourse-power framework (1977, 1980) and Bucholtz and Hall's sociocultural linguistic model of identity (2005) to situate the findings within broader structures of institutional control and community resistance.

## FINDINGS AND DISCUSSION

## FINDNGS

### Forms of Linguistic Innovation

#### *Phonetic Modifications*

Users across all platforms frequently employed phonetic alterations that preserve semantic meaning while evading algorithmic detection. Table 2 catalogues the coded instances of this strategy, each assigned a unique identifier that encodes the platform, case context, and sequence number.

Table 2: Phonetic Modification Strategies

| Code | Platform | Context | Censored Term | Innovation | Strategy Type | Note |
|---|---|---|---|---|---|---|
| TW-HK-01 | Twitter | Hong Kong | Xi Jinping | "Winnie the Pooh" | Homophonic/cultural substitution | Requires knowledge of internet meme to decode |
| TW-HK-02 | Twitter | Hong Kong | CCP | "see see pee" | Homophonic respelling | Algorithmic evasion |
| TT-LG-01 | TikTok | LGBTQ+ | Transgender | "Tr@ns" | Symbol-phonetic hybrid | Algorithmic evasion |
| FB-CV-01 | Facebook | COVID-19 | Vaccine | "V@cc1ne" | Leetspeak substitution | Algorithmic evasion |
| RD-CV-01 | Reddit | COVID-19 | Ivermectin | "horse paste" | Euphemistic phonetic shift | Ironic community reference |
| TW-HK-03 | Twitter | Hong Kong | Hong Kong | "Heung Gong" | Phonetic romanisation | Identity and evasion |

Phonetic modifications were the most widely distributed strategy across all four platforms and all three case contexts. Twitter (TW-HK-01, TW-HK-02, TW-HK-03) showed the highest frequency of homophonic substitution in political

discourse, while TikTok (TT-LG-01) and Facebook (FB-CV-01) predominantly employed leetspeak variants for identity-related and health-related terminology respectively. The substitution recorded as TW-HK-01 — replacing Xi Jinping with a reference to an internet meme figure — represents the most socially complex instance of homophonic substitution in this category, requiring cultural knowledge of the meme for decoding. Reddit's single phonetic instance (RD-CV-01) adopts an ironic register absent from other platforms.

**Symbolic and Visual Substitutions**

Emoji and typographic characters functioned as alternative signification systems, exploiting the multimodal affordances of digital platforms. Table 3 presents all documented instances of this strategy alongside their assigned codes.

Table 3: Symbolic and Visual Substitution Strategies

| Code | Platform | Context | Censored Term | Innovation | Strategy Type | Note |
|---|---|---|---|---|---|---|
| TT-LG-02 | TikTok | LGBTQ+ | Gay | "🌈 fam" | Emoji substitution | Identity marker |
| TT-LG-03 | TikTok | LGBTQ+ | Lesbian | "L e s b i a n" | Spacing modification | Algorithmic evasion |
| TW-HK-04 | Twitter | Hong Kong | Police | "🐷" | Emoji substitution | Political critique |
| FB-CV-02 | Facebook | COVID-19 | COVID-19 | "C💉V💉D" | Symbol insertion | Evasion and ideological encoding |
| TW-HK-05 | Twitter | Hong Kong | Tiananmen (June 4th) | "35th of May" | Visual/calendar displacement | Historical memory preservation |
| TT-LG-04 | TikTok | LGBTQ+ | Trans | "🏳️ community" | Emoji flag substitution | Identity and community signal |
| RD-CV-02 | Reddit | COVID-19 | Vaccinated | "💉'd" | Symbol-verb fusion | Ingroup/outgroup marker |

TikTok recorded the highest concentration of symbolic substitutions (TT-LG-02, TT-LG-03, TT-LG-04), consistent with its video-centred, multimodal platform affordances that normalise emoji as primary communicative units. The calendar displacement strategy documented as TW-HK-05 is the most culturally layered instance in this category, requiring knowledge of Chinese political history to decode: it encodes a specific anniversary through a date that is calendrically impossible, comprehensible only to those familiar with the significance of early June in modern Chinese history. The Facebook innovation coded FB-CV-02 demonstrates how symbolic insertions simultaneously disrupt keyword detection and encode ideological positioning within the modified form itself.

**Semantic Obfuscation**

Users developed community-internal coded language systems requiring shared cultural knowledge for interpretation. These were most densely concentrated on Reddit and TikTok, as shown in Table 4.

Table 4: Semantic Obfuscation Strategies

| Code | Platform | Context | Censored Term | Innovation | Strategy Type | Note |
|---|---|---|---|---|---|---|
| TT-LG-05 | TikTok | LGBTQ+ | Gay | "spicy straight" | Euphemistic substitution | Ingroup humour and evasion |
| TW-CV-01 | Twitter | COVID-19 | Suicide / killed | "unalived" | Euphemistic substitution | Platform policy compliance |
| RD-CV-03 | Reddit | COVID-19 | Unvaccinated | "PureBloods" | Ingroup identity coinage | Anti-vaccination identity construction |
| RD-CV-04 | Reddit | COVID-19 | Vaccinated | "Vaxxed" | Ingroup/outgroup marker | Social boundary |

| Code | Platform | Context | Censored Term | Innovation | Strategy Type | Note |
|---|---|---|---|---|---|---|
| | | | | | | demarcation |
| TW-HK-06 | Twitter | Hong Kong | Protest | "going for a walk" | Metaphorical displacement | Event coordination |
| RD-LG-01 | Reddit | LGBTQ+ | Community-specific terms | Thread-specific vernacular | Sustained ingroup slang system | Long-term community vocabulary |
| TW-HK-07 | Twitter | Hong Kong | Arrested | "having tea with the authorities" | Ironic euphemism | Community reporting |
| FB-CV-03 | Facebook | COVID-19 | Lockdown restrictions | "freedom measures" | Semantic inversion | Ideological reframing |

Reddit produced the most elaborate and sustained coded language systems (RD-CV-03, RD-CV-04, RD-LG-01), reflecting its threaded forum structure which enables long-term community vocabulary development. The term recorded as TW-CV-01 — a euphemistic verb substitute for suicide-related language originating on TikTok — appeared subsequently across Twitter and Reddit within the data collection window, indicating cross-platform lexical diffusion. The binary vocabulary documented as RD-CV-03 and RD-CV-04 constructed a symmetrical ingroup/outgroup system that organised community social relations around vaccination status as an identity category, with one term (RD-CV-03) drawing on a literary allusion to encode ideology within an apparently innocuous cultural reference.

**Hashtag Innovation**

Hashtag manipulation was identified on Twitter and TikTok, where hashtags function as the primary content discovery and community organisation mechanism. Table 5 presents the full set of documented innovations in this category.

Table 5: Hashtag Innovation Strategies

| Code | Platform | Context | Original Hashtag | Innovation | Strategy Type | Note |
|---|---|---|---|---|---|---|
| TT-LG-06 | TikTok | LGBTQ+ | #Pride | "#Pride2k21" | Temporal variation | Content discoverability |
| TW-CV-02 | Twitter | COVID-19 | #Lockdown | "#Fr33d0m" | Leetspeak modification | Ideological reframing |
| TW-CV-03 | Twitter | COVID-19 | #Pandemic | "#Plandemic" | Orthographic/semantic fusion | Conspiracy framing |
| TW-HK-08 | Twitter | Hong Kong | #HongKong | Fragmented nested variants | Nested tag fragmentation | Algorithmic evasion |
| TT-LG-07 | TikTok | LGBTQ+ | #Lesbian | "#LesbianVisibility2021" | Date-appended expansion | Community organisation |
| TW-CV-04 | Twitter | COVID-19 | #Vaccine | "#Vaxx" | Orthographic truncation | Discoverability maintenance |
| TT-LG-08 | TikTok | LGBTQ+ | #Trans | "#TransIsBeautiful22" | Affirmative temporal variation | Identity affirmation and evasion |

Hashtag innovations on TikTok (TT-LG-06, TT-LG-07, TT-LG-08) were predominantly temporal in character — appending dates or year variants to generate new, unblocked tag strings while preserving community discoverability. On Twitter, hashtag modification more frequently fused orthographic alteration with semantic reframing: the instance coded TW-CV-03 simultaneously evades keyword detection

and encodes a conspiratorial interpretation of the pandemic within the modified form itself, making it both a linguistic evasion strategy and an ideological statement. The truncation strategy recorded as TW-CV-04 prioritises discoverability maintenance over semantic modification.

## Case Study Analyses

### Hong Kong Protest Movement (2019–2021)

Table 6 presents the complete set of innovations documented within the Hong Kong protest corpus, drawn exclusively from Twitter (TW-HK-01 through TW-HK-08) and Reddit (RD-HK-01).

Table 6: Hong Kong Protest Movement — Censored Terms and Innovations

| Code | Platform | Censored Term | Innovation | Strategy Type | Function |
|---|---|---|---|---|---|
| TW-HK-01 | Twitter | Xi Jinping | "Winnie the Pooh" | Homophonic/cultural substitution | Political ridicule and evasion |
| TW-HK-02 | Twitter | CCP | "see see pee" | Homophonic respelling | Algorithmic evasion |
| TW-HK-03 | Twitter | Hong Kong | "HK" / "Heung Gong" | Acronym / phonetic romanisation | Identity and evasion |
| TW-HK-04 | Twitter | Police | " 🐷 " | Emoji substitution | Political critique |
| TW-HK-05 | Twitter | Tiananmen (June 4th) | "35th of May" | Calendar displacement | Historical memory preservation |
| TW-HK-06 | Twitter | Protest | "going for a walk" | Metaphorical displacement | Event coordination |

| Code | Platform | Censored Term | Innovation | Strategy Type | Function |
|---|---|---|---|---|---|
| TW-HK-07 | Twitter | Arrested | "having tea with the authorities" | Ironic euphemism | Community reporting |
| RD-HK-01 | Reddit | Pro-democracy movement | "the situation" | Vague nominal displacement | Risk reduction |

The Hong Kong corpus revealed the highest degree of culturally embedded evasion strategies in the dataset. Seven of the eight innovations presented in Table 6 require knowledge of Chinese political history, internet meme culture, or Cantonese linguistic context to decode — a higher threshold of cultural competence than is required for any other case context in this study. The datum coded TW-HK-05 is the most technically sophisticated in the entire corpus: it encodes a specific historical date through a calendrical impossibility that is meaningful only to those who know that May has 31 days and that the date in question falls four days into the following month. The Reddit instance (RD-HK-01) adopts a markedly different evasion logic, favouring vague nominal displacement over culturally specific coding in a manner consistent with Reddit's more geographically diffuse user base.

**LGBTQ+ Censorship on TikTok**

Table 7 catalogues the innovations documented within the LGBTQ+ TikTok corpus, all of which were recovered from TikTok (TT-LG-01 through TT-LG-08), with supplementary Reddit data recorded separately as RD-LG-01.

Table 7: LGBTQ+ Content — Censored Terms and Innovations

| Code | Platform | Censored Term | Innovation | Strategy Type | Function |
|---|---|---|---|---|---|
| TT-LG-01 | TikTok | Transgender | "Tr@ns" | Symbol-phonetic hybrid | Algorithmic evasion |

| Code | Platform | Censored Term | Innovation | Strategy Type | Function |
|---|---|---|---|---|---|
| TT-LG-02 | TikTok | Gay | " 🌈 fam" | Emoji substitution | Identity marker |
| TT-LG-03 | TikTok | Lesbian | "L e s b i a n" | Spacing modification | Algorithmic evasion |
| TT-LG-04 | TikTok | Trans | " 🏳️‍⚧️ community" | Emoji flag substitution | Identity and community signal |
| TT-LG-05 | TikTok | Gay | "spicy straight" | Euphemistic substitution | Ingroup humour and evasion |
| TT-LG-06 | TikTok | #Pride | "#Pride2k21" | Temporal hashtag variation | Content discoverability |
| TT-LG-07 | TikTok | #Lesbian | "#LesbianVisibility2021" | Date-appended expansion | Community organisation |
| TT-LG-08 | TikTok | #Trans | "#TransIsBeautiful22" | Affirmative temporal variation | Identity affirmation and evasion |

The LGBTQ+ TikTok corpus produced the greatest variety of strategy types within a single platform, encompassing phonetic (TT-LG-01), symbolic (TT-LG-02, TT-LG-04), spacing (TT-LG-03), euphemistic (TT-LG-05), and hashtag innovations (TT-LG-06, TT-LG-07, TT-LG-08). Notably, several of the innovations documented in this corpus serve a dual function: they evade algorithmic detection while simultaneously operating as identity affirmation statements. The instances coded TT-LG-08 and TT-LG-04, for example, are not merely evasive — they are assertive, transforming the necessity of linguistic modification into an opportunity for visible identity expression. This dual function distinguishes LGBTQ+ linguistic innovation from the predominantly evasion-focused strategies observed in the political and health discourse contexts.

**COVID-19 Discourse on Twitter, Facebook, and Reddit**

Table 8 presents the COVID-19 corpus, distributed across Twitter (TW-CV-01 through TW-CV-04), Facebook (FB-CV-01 through FB-CV-03), and Reddit (RD-CV-01 through RD-CV-04), making it the only case context represented on three platforms.

Table 8: COVID-19 Discourse — Censored Terms and Innovations

| Code | Platform | Censored Term | Innovation | Strategy Type | Function | Context |
|---|---|---|---|---|---|---|
| FB-CV-01 | Facebook | Vaccine | "V@cc1ne" | Leetspeak substitution | Algorithmic evasion | Anti-vaccination |
| FB-CV-02 | Facebook | COVID-19 | "C💉V💉D" | Symbol insertion | Evasion + ideological encoding | Anti-vaccination |
| FB-CV-03 | Facebook | Lockdown restrictions | "freedom measures" | Semantic inversion | Ideological reframing | Anti-restriction |
| TW-CV-01 | Twitter | Suicide / killed | "unalived" | Euphemistic substitution | Platform policy compliance | Content moderation |
| TW-CV-02 | Twitter | #Lockdown | "#Fr33d0m" | Leetspeak modification | Ideological reframing | Anti-restriction |
| TW-CV-03 | Twitter | #Pandemic | "#Plandemic" | Orthographic/semantic fusion | Conspiracy framing | Anti-vaccination |

| Code | Platform | Censored Term | Innovation | Strategy Type | Function | Context |
|---|---|---|---|---|---|---|
| TW-CV-04 | Twitter | #Vaccine | "#Vaxx" | Orthographic truncation | Discoverability maintenance | Pro/Anti-vaccination |
| RD-CV-01 | Reddit | Ivermectin | "horse paste" | Euphemistic phonetic shift | Ironic community reference | Pro-vaccination |
| RD-CV-02 | Reddit | Vaccinated | "💉'd" | Symbol-verb fusion | Ingroup/outgroup marker | Anti-vaccination |
| RD-CV-03 | Reddit | Unvaccinated | "PureBloods" | Ingroup identity coinage | Community identity construction | Anti-vaccination |
| RD-CV-04 | Reddit | Vaccinated | "Vaxxed" | Ingroup/outgroup marker | Social boundary demarcation | Both sides |

The COVID-19 corpus is the largest of the three case contexts and the only one distributed across three platforms, allowing for meaningful cross-platform comparison within a single case context. Facebook innovations (FB-CV-01, FB-CV-02, FB-CV-03) were predominantly symbol-insertion strategies, consistent with a platform user base less familiar with leetspeak conventions than Reddit's. The Reddit instances present the most ideologically charged innovations in this corpus: the binary coded RD-CV-03 and RD-CV-04 constructed a symmetrical identity vocabulary organised around vaccination status, while RD-CV-01 adopts an ironic register that signals pro-vaccination positioning through mock-derogatory terminology. The Twitter datum TW-CV-01 is notably distinct from the rest of the COVID-19 corpus in that its evasive function targets platform content moderation policy rather than political censorship.

**Cross-Platform Summary**

Tables 9 and 10 provide aggregate distributions of innovation strategies across platforms and case contexts respectively, drawing together all coded instances documented in Tables 2 through 8.

Table 9: Distribution of Innovation Strategies Across Platforms

| Strategy Type | Twitter | TikTok | Facebook | Reddit | Total |
|---|---|---|---|---|---|
| Phonetic modification | 3 | 1 | 1 | 1 | 6 |
| Symbolic/visual substitution | 2 | 3 | 2 | 1 | 8 |
| Semantic obfuscation | 2 | 1 | 1 | 4 | 8 |
| Hashtag innovation | 4 | 4 | 0 | 0 | 8 |

Table 10: Distribution of Innovation Strategies Across Case Contexts

| Strategy Type | Hong Kong (HK) | LGBTQ+ (LG) | COVID-19 (CV) | Total |
|---|---|---|---|---|
| Phonetic modification | 3 | 1 | 2 | 6 |
| Symbolic/visual substitution | 2 | 4 | 2 | 8 |
| Semantic obfuscation | 3 | 2 | 3 | 8 |
| Hashtag innovation | 1 | 3 | 4 | 8 |
| **Total per context** | 9 | 10 | 11 | 30 |

Twitter produced the broadest range of strategy types across all contexts and was the only platform represented in all three case contexts. TikTok dominated symbolic and hashtag innovations, reflecting its multimodal, video-centred affordances and younger user base. Reddit produced the most elaborate semantic obfuscation systems, consistent with its sustained community forum structure. Facebook showed the narrowest innovation range, relying primarily on symbol

insertion and phonetic modification. As Table 10 shows, the COVID-19 context generated the largest number of documented innovations overall (11), followed closely by LGBTQ+ content (10) and Hong Kong political discourse (9). Cross-platform lexical diffusion was observed across all four platforms, most clearly evidenced by the migration of the TT-LG-05 / TW-CV-01 euphemistic verb from TikTok to Twitter and Reddit within the data collection window.

Table 11: Empirical–Theoretical Mapping of Digital Censorship, Identity, and Linguistic Innovation

| Subheading | Core Theoretical Concept | Empirical Observation (Data Pattern) | Illustrative Examples from Digital Discourse | Analytical Interpretation | Key Scholars |
|---|---|---|---|---|---|
| **Foucauldian Configurations of Power, Discourse, and Resistance** | Power as productive; discourse proliferation; dispositif; subjugated knowledges | Iterative cycle: censorship → linguistic innovation → platform adaptation → renewed innovation | "unalived" (for "killed"), "seggs" (for "sex"), creative misspellings, emoji substitutions | Censorship generates rather than suppresses discourse; users operate tactically within platform constraints, producing alternative semiotic systems | Michel Foucault; Michel de Certeau, (1980) |
| **Identity as Sociolinguistic Practice in Digitally Mediated Contexts** | Identity as emergent, indexical, relational | Linguistic choices signal group membership, awareness of censorship, and ideological stance | Use of coded language to signal "in-group" status; avoidance of platform-trigger words | Identity is not pre-existing but performed through linguistic adaptation; language becomes both communicative and | Mary Bucholtz; Kira Hall, (2005) |

| | | | | | |
|---|---|---|---|---|---|
| | | | | identity-indexing resource | |
| **(Emergence dimension)** | Identity constructed in interaction | Real-time creation of new lexical forms in response to moderation | Sudden viral adoption of euphemisms during platform crackdowns | Identity emerges situationally through participation in evolving discourse practices | Bucholtz & Hall (2005) |
| **(Indexicality dimension)** | Linguistic forms carry social meanings beyond reference | Marked forms index resistance, awareness, and digital literacy | Using "unalived" signals awareness of platform censorship norms | Words acquire ideological weight; linguistic form becomes meta-communicative | Michael Silverstein, (2003) |
| **(Relationality dimension)** | Identity through adequation and distinction | Clear boundary between users who "get" coded language and those who do not | Insider vs outsider comprehension of euphemisms | Language functions as boundary-making tool, reinforcing community cohesion | Bucholtz & Hall (2005) |
| **Linguistic Innovation and Language Change in Digital Ecologies** | Language change through variation, diffusion, and innovation | Stabilisation and spread of censorship-driven lexical items beyond original contexts | "unalived" used in mainstream discourse, not only censored contexts | Tactical innovations evolve into systemic linguistic change; digital platforms accelerate diffusion | William Labov, (1994) |
| **Digital Public Spheres, Epistemic** | Public sphere; communicative | Emergence of semiotically opaque | Encoded discussions that require community | Censorship produces stratified communicat | Jürgen Habermas (1984) |

| | | | | | |
|---|---|---|---|---|---|
| **Access, and Democratic Discourse** | accessibility; epistemic fragmentation | discourse inaccessible to outsiders | knowledge to interpret | ive spaces, limiting universal accessibility and deliberation | |
| **Censorship and Institutional Opacity** | Governance vs legibility paradox | Moderation systems fail to detect encoded harmful content | Harmful speech disguised through euphemism evades detection | Censorship reduces transparency and effectiveness of regulation mechanisms | Platform governance scholarship |
| **Censorship, Collective Identity, and Digital Subcultures** | Communities of practice; social identity; linguistic capital | Shared coded language becomes marker of belonging and symbolic value | Group-specific slang used consistently within subcultures | Linguistic competence becomes social capital; reinforces solidarity and resistance identity | Etienne Wenger (1998); Henri Tajfel & Turner C John (1979); Pierre Bourdieu (1991) |
| **Language, Power, and Social Cohesion** | Language as symbolic power and social structuring force | Persistent use of alternative lexicons across communities | Stable subcultural linguistic norms | Language mediates both resistance to power and internal cohesion of groups | Basil Bernstein (1971) |

## DISCUSSION

The findings of this study substantiate the proposition that censorship and linguistic innovation are bound in a recursive, dialectical relationship. Attempts by digital platforms to suppress particular forms of expression do not culminate in discursive absence; rather, they engender alternative semiotic strategies through which users actively reconstitute meaning. When specific lexical items are suppressed, users respond with semantically equivalent substitutions, orthographic

distortions, code-switching practices, and multimodal signification — confirming that censorship operates not merely as a repressive mechanism but as a generative force that stimulates novel linguistic forms. This finding directly addresses the study's primary research objective concerning the relationship between platform governance and linguistic creativity within digitally mediated Nigerian discourse communities. With respect to the guiding research questions, the data demonstrate, first, that users do not abandon communicative intentions under platform restrictions but recalibrate the semiotic vehicles through which those intentions are expressed; second, that such adaptive strategies constitute meaningful forms of agency, positioning users as active participants in the negotiation of communicative norms rather than passive recipients of regulatory control; and third, that censorship-evasive language frequently functions as a marker of in-group solidarity, particularly within marginalised communities where coded expression carries both tactical and symbolic weight.

The findings are broadly consonant with prior research on language, power, and digital communication while simultaneously extending existing frameworks. The observation that platform censorship generates rather than extinguishes linguistic creativity resonates with empirical work on Chinese digital discourse, particularly studies documenting homophonic substitutions and visual puns on Weibo in response to keyword-based filtering (King, Pan, & Roberts, 2013; Roberts, 2018. The study's identification of iterative evasion dynamics, wherein moderation systems and user strategies co-evolve continuously corresponds with platform governance scholarship documenting the inadequacy of static keyword-based classifiers (Wiegand, Ruppenhofer, & Kleinbauer, 2019). Foucault's (1977) understanding of power as productive rather than merely prohibitive is confirmed: censorship systems catalyse the very creativity they seek to suppress. This is equally consistent with Scott's (1990) conceptualisation of hidden transcripts, wherein subordinate groups encode resistance within ostensibly compliant discursive forms a logic the present data demonstrate is not confined to face-to-face interaction but is reproduced and intensified within algorithmically governed communicative

environments. However, the findings complicate the deterministic strand of surveillance capitalism scholarship associated with Zuboff (2019): users demonstrate a persistent and sophisticated capacity for semiotic innovation that resists any framework that construes them primarily as subjects of behavioural manipulation, arguing instead for more dialectical accounts of digital power that hold platform dominance and user agency in productive theoretical tension. An unexpected finding concerns the affective and identitarian dimensions of linguistic evasion: for marginalised communities, coded language functions not merely as tactical evasion but as an expression of collective identity and ideological alignment what Halliday (1978) terms anti-language suggesting that future research should attend more systematically to the affective economies of digital linguistic innovation.

The theoretical implications of these findings span digital sociolinguistics, platform governance studies, and the sociology of communication. The study affirms that meaning is not fixed in discrete lexical tokens but emerges through context, pragmatics, and shared interpretive frameworks, a view that renders structurally inadequate any censorship model that privileges surface-level linguistic features. The findings contribute to scholarship treating online communication as a generative arena for linguistic change (Crystal, 2011; Thurlow & Mroczek, 2011) and, more specifically, suggest that platform governance itself functions as a sociolinguistic variable shaping the trajectories of lexical and semiotic change in ways comparable to, though distinct from, traditional social pressures. Drawing on Butler's (1990) theory of performativity, the study proposes that censorship-evasive strategies should be understood not merely as tactical accommodations but as constitutive dimensions of digital identity: the repeated deployment of coded language embeds resistance within the very structure of selfhood. Goffman's (1959) dramaturgical framework further illuminates the performative adjustments users make under algorithmic surveillance, the digital interface functioning as a regulated front stage, though the present data add that such calibration is not uniformly

experienced as constraining but constitutes, for many users, a source of social capital and communal affiliation.

The practical implications bear directly on platform design and digital governance, particularly in multilingual, postcolonial, and Global South contexts. The persistent failure of keyword-based moderation in the face of user innovation underscores the inadequacy of purely technical approaches to content governance (Gillespie, 2018). In the Nigerian context specifically, the extraordinary linguistic diversity of digital discourse communities characterised by multilingualism, code-switching, and the creative blending of indigenous and metropolitan communicative repertoires demands that moderation infrastructures be developed in close consultation with those communities rather than imposed through generically trained systems calibrated to dominant language varieties, lending empirical support to calls for the decolonisation of digital governance (Birhane, 2020; Mohamed, Png, & Isaac, 2020). The study further cautions against intensified moderation as an unambiguous good, identifying risks of platform migration, fragmentation of digital public spheres, and the formation of increasingly insular discourse communities. Empirically, the study provides systematic documentation of censorship-evasion strategies within a Nigerian digital discourse context, a setting underrepresented in the international literature, thereby expanding the geographic scope of platform governance research beyond predominantly Western and East Asian contexts. Theoretically, it advances an integrated analytical model synthesising Foucauldian discourse theory, Scottian resistance studies, Butlerian performativity, Hallidayan sociolinguistics, and Zuboffian surveillance capitalism scholarship. Normatively, it foregrounds the capacity of linguistically marginalised communities to act, create, and resist within regulatory architectures, participating in the broader scholarly project of theorising the creative and political potentialities of subaltern digital practice. The interplay between control and creativity thus emerges not merely as a technical problem of platform governance but as a defining feature of contemporary communicative life demanding sustained, interdisciplinary scholarly attention.

## CONCLUSION

This study has examined the dialectical relationship between digital censorship, linguistic innovation, and identity construction across four major social media platforms, grounded in Foucault's (1977, 1980) discourse-power framework and Bucholtz and Hall's (2005) sociocultural linguistic model of identity. The findings collectively demonstrate that censorship operates as a generative rather than suppressive force, that censorship-evasive strategies function simultaneously as instruments of circumvention and as constitutive acts of identity affirmation, and that digitally induced linguistic innovation diffuses across platforms at a pace that existing sociolinguistic frameworks, developed primarily in face-to-face contexts, are insufficient to characterise.

Theoretically, the study advances Foucauldian analysis into the domain of platform governance, extends Bucholtz and Hall's identity framework to account for the specific pressures that algorithmic constraint exerts on digital self-constitution, and introduces the non-human moderation system as a qualitatively distinct interlocutor against which users design language — a dimension untheorised in prior applications of Audience Design (Bell, 1984) and Accommodation Theory (Giles, 1973). Practically, the evidence suggests that keyword-based content moderation is both technically ineffective and potentially counterproductive, with governance frameworks attuned to linguistic diversity and communicative freedom preferable to blunt algorithmic lexical enforcement (Gillespie, 2018).

Future research should pursue longitudinal quantitative tracking of cross-platform lexical diffusion, empirical investigation of generative artificial intelligence as an emerging tool of automated linguistic evasion, and extension of this analytical framework to Global South digital contexts, particularly the multilingual environments of West African social media, where postcolonial governance histories and indigenous linguistic resources interact with platform regulation in ways that current scholarship has yet to engage systematically. This study ultimately affirms that language constitutes an irreducibly dynamic site of power, resistance, and identity in digitally mediated life, and that sustained scholarly

engagement with the linguistic consequences of platform governance is both an academic imperative and a contribution to defending communicative freedom in an increasingly regulated digital world.

**REFERENCES**

Androutsopoulos, J. (2006). Multilingualism, diaspora, and the Internet: Codes and identities on German-based diaspora websites. *Journal of Sociolinguistics, 10*(4), 520–547.

Baron, N. S. (2008). *Always on: Language in an online and mobile world*. Oxford University Press.

Baym, N. K. (2015). *Personal connections in the digital age* (2nd ed.). Polity Press.

Bell, A. (1984). Language style as audience design. *Language in Society, 13*(2), 145–204.

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity Press.

Bernstein, B. (1971). *Class, codes and control: Vol. 1. Theoretical studies towards a sociology of language*. Routledge & Kegan Paul.

Birhane, A. (2020). Algorithmic colonization of Africa. *SCRIPTed: A Journal of Law, Technology and Society*, *17*(2), 389–409. https://doi.org/10.2966/scrip.170220.389

Bonilla, Y., & Rosa, J. (2015). #Ferguson: Digital protest, hashtag ethnography, and the racial politics of social media in the United States. *American Ethnologist, 42*(1), 4–17.

Bourdieu, P. (1991). *Language and symbolic power* (G. Raymond & M. Adamson, Trans.; J. B. Thompson, Ed.). Harvard University Press.

Bou-Franch, P., & Garcés-Conejos Blitvich, P. (2014a). Conflict management in massive polylogues: A case study from YouTube. *Journal of Pragmatics*, *73*, 19–36. Academia.edu

_____. (2014b). Gender ideology and social identity processes in online language aggression against women. *Journal of Language Aggression and Conflict*, *2*(2), 226–248. https://doi.org/10.1075/jlac.2.2.03bou

boyd, d. (2014). *It's complicated: The social lives of networked teens*. Yale University Press.

boyd, d., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication, 13*(1), 210–230.

Butler, J. (1990). *Gender trouble: Feminism and the subversion of identity.* Routledge

Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford University Press.

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

Cover, R. (2016). *Digital identities: Creating and communicating the online self*. Academic Press.

Crystal, D. (2011). *Internet linguistics: A student guide*. Routledge.

Daniels, J. (2020). The algorithmic rise of the "alt-right." *Social Media + Society, 6*(2), 1–12.

de Certeau, M. (1984). *The practice of everyday life* (S. Rendall, Trans.). University of California Press.

Deibert, R., Palfrey, J., Rohozinski, R., & Zittrain, J. (Eds.). (2008). *Access denied: The practice and policy of global Internet filtering*. MIT Press.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Routledge.

Fakhira, R., & Sartini, N. W. (2024). Cross-cultural pragmatic study of complaint strategies in Banjarese and Javanese. *LET: Linguistics, Literature and English Teaching Journal, 14*(2), 270–290. https://jurnal.uin-antasari.ac.id/index.php/let/article/view/13764

Flew, T. (2021). *Regulating platforms*. Polity Press.

Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). Pantheon Books.

Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings, 1972–1977* (C. Gordon, Ed.; C. Gordon, L. Marshall, J. Mepham, & K. Soper, Trans.). Pantheon Books.

Foucault, M. (1984). *The history of sexuality, Volume 1: The will to knowledge*. Penguin.

franzke, a. s., Bechmann, A., Zimmer, M., Ess, C., & the Association of Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0*. https://aoir.org/reports/ethics3.pdf

Gee, J. P. (2014). *An introduction to discourse analysis: Theory and method* (4th ed.). Routledge.

Gerrard, Y. (2018). Beyond the hashtag: Circumventing content moderation on social media. *New Media & Society, 20*(12), 4492–4511.

Giles, H. (1973). Accent mobility: A model and some data. *Anthropological Linguistics, 15*(2), 87–105.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Goffman, E. (1959). *The presentation of self in everyday life.* Doubleday Anchor Books.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.

Habermas, J. (1984). *The theory of communicative action, Vol. 1: Reason and the rationalization of society* (T. McCarthy, Trans.). Beacon Press.

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning.* Edward Arnold.

Herring, S. C. (2007). A faceted classification scheme for computer-mediated discourse. *Language@Internet*, *4*(1). https://www.languageatinternet.org/articles/2007/761

Jones, R. H., Chik, A., & Hafner, C. A. (Eds.). (2015). *Discourse and digital practices: Doing discourse analysis in the digital age*. Routledge.

King, G., Pan, J., & Roberts, M. E. (2013). How censorship in China allows government criticism but silences collective expression. *American Political Science Review*, *107*(2), 326–343.

Labov, W. (1994). *Principles of linguistic change: Vol. 1. Internal factors*. Blackwell.

Labov, W. (1972). *Sociolinguistic patterns*. University of Pennsylvania Press.

MacKinnon, R. (2012). *Consent of the networked: The worldwide struggle for Internet freedom*. Basic Books.

Marwick, A., & boyd, d. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society, 13*(1), 114–133.

Marwick, A., & boyd, d. (2018). Understanding privacy at the margins. *International Journal of Communication, 12*, 1157–1165.

McCulloch, G. (2019). *Because Internet: Understanding the new rules of language*. Riverhead Books.

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, *33*(4), 659–684. https://doi.org/10.1007/s13347-020-00405-8

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

Roberts, S. T. (2019). *Behind the screen: Content moderation in the shadows of social media*. Yale University Press.

Roberts, M. E. (2018). *Censored: Distraction and diversion inside China's Great Firewall.* Princeton University Press.

Seargeant, P., & Tagg, C. (Eds.). (2014). *The language of social media: Identity and community on the internet.* Palgrave Macmillan. https://doi.org/10.1057/9781137029317

Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication, 23*(3–4), 193–229. https://doi.org/10.1016/S0271-5309(03)00013-2

Suzor, N. (2019). *Lawless: The secret rules that govern our digital lives*. Cambridge University Press.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Brooks/Cole.

Thurlow, C., & Mroczek, K. (Eds.). (2011). *Digital discourse: Language in the new media*. Oxford University Press.

Tufekci, Z. (2017). *Twitter and tear gas: The power and fragility of networked protest*. Yale University Press.

Turkle, S. (2015). *Reclaiming conversation: The power of talk in a digital age*. Penguin Press.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge University Press. https://doi.org/10.1017/CBO9780511803932

Wiegand, M., Ruppenhofer, J., & Kleinbauer, T. (2019). Detection of abusive language: The problem of biased datasets. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for*

*Computational Linguistics: Human Language Technologies* (Vol. 1, pp. 602–608). Association for Computational Linguistics.

Yus, F. (2023). Pragmatics and the Internet. In The Encyclopedia of Applied Linguistics (Pragmatics), C.A.Chapelle, N.Taguchi & D.Kadar (Eds.). https://doi.org/10.1002/9781405198431.wbeal0309.pub2.

Zhou, J., Zhu, Y., & Jiang, M. (2021). Online resistance: The language games of censorship evasion in China. *New Media & Society, 23*(7), 1768–1785.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* PublicAffairs.

Zuckerman, E. (2010). *Rewire: Digital cosmopolitans in the age of connection*. W. W. Norton & Company.

Zuckerman, E. (2019). The case for digital public infrastructure. Knight First Amendment Institute.