



TAMSAAL

ISSN Print: [3007-5939](#)

ISSN Online: [3007-5947](#)

Volume 4, Number 9, 2026, Pages 483 – 502

Journal Home Page

<https://tamsaal.pk/index.php/tamsaal/index>

TAMSAAL
JOURNAL



Coded Conversations: A Critical Discourse Analysis of Algospeak as a Subversive Linguistic Practice on X

Aneeza Zara¹

¹MPhil. English Linguistics Scholar NUML, Rawalpindi. Email: aneezaraqureshi10@gmail.com

ARTICLE INFO

Article History:

Received: August 16, 2026
Revised: September 08, 2026
Accepted: September 25, 2026
Available Online: September 27, 2026

Keywords:

algospeak, critical discourse analysis, algorithmic content moderation, platform governance, digital resistance, X

Corresponding Author:

Aneeza Zara

Email:

aneezaraqureshi10@gmail.com

DOI: <https://doi.org/10.59075/tamsaal.v4i9.223>

ABSTRACT

Platforms decide which words are too risky to publish, and users learn to speak around them. This study examines algospeak, the deliberate misspelling, symbol substitution, and re-coding of words to evade automated moderation, as a subversive linguistic practice on X (formerly Twitter). Using Critical Discourse Analysis (Wodak & Meyer, 2022) and platform studies scholarship (Gorwa et al., 2020), it analyzes 25 English-language posts from January 2025 to May 2026, sampled purposively across mental health discourse, political speech, meta-discourse on moderation, and emoji-based encoding. Six strategies emerged: lexical substitution, phonetic distortion, graphic and numeric distortion, acronymic encoding, subcultural lexicons, and emoji displacement. The findings show that the burden of encoding falls unevenly on users discussing mental health, political marginality, and geopolitically sensitive topics, and that many users treat algospeak as a conscious act of resistance. The study recommends context-sensitive moderation, greater transparency, and algospeak literacy in digital education.



© 2026 The Authors, Published by ULLR. This is an Open Access Article under the Creative Common Attribution ([CC-BY 4.0](#))

Introduction

Background and Context

Language is always changing in relation to power. But, like most things, languages of marginalized groups have developed code-switches that are used beneath the radar of dominant institutional power, from the coded argots of persecuted minorities to the slang of politically surveilled communities. The digital age has created its own brand of this perennial phenomenon which is more visible, more ubiquitous and more technically mediated than its historical counterparts. On X (formerly Twitter), a new twist on the old concept of “language”, algospeak, and its implications

for linguistic creativity, digital censorship, platform governance, and the continuing struggle between algorithmic power and user agency have arisen in response to the platform's automated content moderation systems.

Algospeak is the deliberate misspelling, manipulation or ingenious restatement of words and phrases that are expected to activate automatic moderation systems (Steen & Prah, 2023). In cases where the algorithm may trigger the word "suicide", users can simply write "unalive", but in cases where the word "killed" would potentially be suppressed, the word "k*lled" circulates with relative freedom. Numbers stand for syllables, intentional misspellings circumvent keyword filters, and a collection of emojis can carry a ton of coded meaning outside of what was originally intended. These linguistic innovations are not just a matter of creative style; they're a result of a population's constant interaction with the invisible but powerful logic of platform algorithms (Gorwa et al., 2020). To understand algospeak, more than its formal description is required, it is necessary to critically question the power structure that necessitates such an encoding in the first place.

Statement of the Problem

Although the phenomenon of algospeak has spread like wildfire on the social media, there is still limited attention paid to it in scholarship. The technical architectures, governance implications, and documented impact of algorithmic content moderation have been major areas of focus in existing platform studies (Gorwa et al., 2020). Likewise, Critical Discourse Analysis (CDA) is a well-documented approach that has been used to analyze the relation of power created and reproduced by language (Wodak & Meyer, 2016). The intersection of these two streams, however, of what it means for ordinary users to linguistically negotiate, resist and subvert algorithmic governance (in the process of a practice such as algospeak) has not been subject to a critical discussion.

This is especially relevant in regards to X, a platform that has seen its content review and user governance policies change rapidly and sometimes in an opaque way since it was bought by Elon Musk in 2022. The platform's instability has exacerbated users' doubts about what speech is acceptable and led to a timely and theoretically urgent examination of their adaptive linguistic strategies. The goal of this research is to shed light on the communicative agency exercised or inhibited by users in relation to a particular linguistic phenomenon, but also to offer a glimpse into the discursive context in which this agency is exercised.

Theoretical Framework

The theoretical and methodological framework of this study is based on Critical Discourse Analysis, which views language as a form of social practice that is part of the relations of power and ideology (Wodak and Meyer, 2016). CDA goes beyond content to address issues of who can speak, when, and how, and the consequences of their activity at the institutional level. In this sense, algorithmic moderation systems that are used on X are not technical devices, but rather device of values, of silences, of differential communicative visibility. The framework is supported by platform studies scholarship which places social media platforms as active actors in governance, and not passive actors in infrastructure (Gorwa et al., 2020). Platforms wield power in three ways: by their design, by their moderation policies, and most importantly, by the algorithms they use to amplify, suppress or make certain content invisible. In this double perspective; Algospeak appears

as a counter-discursive practice, a language practice that engages in contestation and deconstruction of the ideological architecture of the language.

Research Questions

The study seeks to respond to the following two research questions:

1. What are the ways in which linguistic strategies constituting algospeak on X function as a form of resistance against algorithmic content moderation?
2. How do algospeak practices on X reflect and reproduce broader power relations between platform governance structures and their user communities

Significance

This research has several contributions. As a theoretical framework, it promises to be applicable to the field of algorithmic government, which offers new opportunities and problems for discourse analysis. The study shows that the methods of critical linguistic inquiry, such as the focus on lexical selection, semantic shift and the ideology of communication, are still very fruitful in the digitally mediated and algorithmically constrained sphere of communication.

The study provides an empirical analysis of the algospeak phenomenon in X at a pivotal moment for the platform's development. Due to the increasing concerns regarding the moderation of social media platforms by regulators, civil society organizations, as well as scholars, it is important to have fine grained analyses of users' reactions to such practices for a full understanding of the digital public sphere (Steen & Prahl, 2023). In practice, the results have implications for platform designers, content moderators, digital rights defenders, and researchers investigating the nature of the relationship between automated systems and communicative creativity of humans.

More generally, it is a study that is inspired by the idea that the study of language under constraint is a fundamental study of power. Users' permitted and disallowed utterances, and the strategies they develop to circumvent those expectations, suggest the shape of the social and institutional contexts in which they function (Ahmed & Gillespie, 2022). The encoded substitutions of algospeak aren't just a linguistic curiosity, they're an adaptive and continuous insistence of the right to communicate, a right that is very human.

Delimitations

The study is delimited in numerous ways. First, the analysis focuses on the textual content posted on X (formerly Twitter) and does not account for other social media platforms, such as TikTok, Instagram, and YouTube, where algospeak can be very different in terms of form and function, depending on the differing architectures of moderation and demographics of users. A comparative study across platforms would be extremely useful, but for the purposes of the present study a single-platform study is warranted. Secondly, this study investigates algospeak only in English. This linguistic restriction is both a result of the overwhelming majority of English-language research on CDA and platform studies and of the analyst's capacity. It is recognized, however, that algospeak exists in several languages and that non-English speakers might have different encoding strategies due to the phonological and morphological characteristics of their language. Third, this study does not aim at assessing the effectiveness of algospeak as a technical tool to circumvent particular moderation tools, nor does it try to measure the result on the platform side, like the rate

of removal or the shadow-banning. It is not technically about its success rate, but about what algospeak signifies, how it's used as a social and linguistic practice, and what it tells us about how power is distributed in its operations on the platform. Last but not least, the study relies on a corpus that has been gathered in a fixed period and not on the entire history of algospeak in X.

Literature Review

Introduction

This chapter summarizes the current research on algospeak as a subversive form of communication on X. Thematic organization of the review takes place in four interconnected domains: emergence and nature of the algospeak as a digital linguistic phenomenon, the architecture and ideologies of algorithmic content moderation, the theorization of digital and platform mediated communication through a Critical Discourse Analysis framework, and theorization of language as a space of contestation of institutional power. These threads of scholarship provide the theoretical basis for the present study and reveal the key lacuna that the study aims to fill.

Algospeak: Emergence and Characteristics

Algospeak as an isolated linguistic phenomenon is a relatively new study, on account of the novelty of its conditions. The euphemism and avoidance strategies that social media users use to encode their speech in anticipation of algorithmic detection and suppression are encapsulated in a single term: algospeak, as described by Steen and Prahl (2023). They uncover a number of common formal strategies, such as vowel substitution, intentional misspelling, numeric substitution, and the semantic extension of emojis, and place them in a wider ecosystem of platform affordances and user adaptations. Importantly, Steen and Prahl (2023) set apart algospeak from previous varieties of internet slang or creative language play, pointing to its highly reactive nature as it is not merely a product of creative innovation, but of institutional pressure from automated mods that make some words and topics “*dangerous to communicate*”.

Algospeak comes not out of the blue, but from a certain power dynamic among users and platforms, and its formal characteristics can be directly influenced by the technical logics of the systems that users try to weaken. Such lexical creativity such as “unalive” for suicide, “legal” for legal in sensitive contexts, and “seggs” for sex, is a result of a population of users that is able to sculpt the contours of algorithmic surveillance and react with adaptive linguistic resourcefulness (Steen & Prahl, 2023). It's not just a linguistic artifact, it's a social act, it's a creative practice and it's a survival strategy which makes algospeak exciting as an object of critical inquiry.

Content Moderation and Platform Power through Algorithms

Algospeak can only be understood through an understanding of the systems of moderation that it is based on. In their thorough overview of algorithmic content moderation, Gorwa et al. (2020) look at both the technical and political hurdles to automating platform governance at scale. They say these automated systems of moderation, which now review content at volumes that are unmanageable for human readers and writers, are not just functional tools they are systems of governance in which there are “*assumptions about what is acceptable speech*.” Establishing a flag for a keyword, suppressing hashtags, or deprioritizing a content type is never a technically neutral

process, but rather a series of implicit and/or explicit value judgments regarding risk, harm, and communicative legitimacy.

The opacity of these systems is also a significant aspect of platform power as they have been pointed out by Gorwa et al. (2020). For users, they are seldom told exactly how their content will be judged and at what point it will be moderated and what the moderation process will be like, or what makes suppressed content eligible for review or reappearance. This opacity isn't accidental, but rather a structural phenomenon that happens to systematically disadvantage users of platforms and help to produce the conditions where practices such as algospeak are needed. If the user is not sure what will be monitored, they must resort to precautionary measures that take the widest possible interpretation of surveillance; algospeak is one of these measures, among others.

The implications of this asymmetry within a governing body are significant. In the analysis of content practices on alternative and fringe platforms, Zeng and Schäfer (2021) illustrate how *moderation-as-response* has become a part of the linguistically-adapting migration process in reaction to platform power. Their research finds that people do not just blindly comply with what is imposed by algorithmic governance, but negotiate, work around and adjust it, sometimes in ways that are admissible in communication, and sometimes in ways that are not. This ambivalence is at the heart of the narrative of algospeak and needs to be kept in mind when analyzing what comes next.

Critical Discourse Analysis and Digital Communication

The method which supports this study is Critical Discourse Analysis. According to Wodak and Meyer (2022), CDA isn't a single method but a group of techniques that have one thing in common: an interest in the interconnection of language, power and ideology. One aspect which sets CDA apart from other kinds of discourse analysis is the critical nature of its approach, which does not assume that language is a neutral communicative medium, but is instead a social practice in which power relations are created, maintained, and most importantly challenged. CDA analysts focus on how certain discursive practices 'naturalize' certain social practices, marginalize certain speakers and close down possibilities for meaning making.

When applied to the digital communication, CDA faces familiar ground and new challenges. The multimodal complexity, volume and velocity of social media discourse call for adaptations of traditional analytical approaches and the algorithmic mediation of digital platforms add a new level of agency that has no direct counterpart in previous discourse settings (Wodak & Meyer, 2022). Platforms do not just serve as a space for discourse; they actually design, select, privilege, and suppress it. Because CDA's focus on the institutional aspects of discourse means that it also offers concepts that are useful for analyzing not just what users say but the structural conditions in which they say it, this approach particularly lends itself to this environment.

Language, Power, and Resistance

The last, and most related, strand of literature to this research is the theorization of language as a space of resistance. In their analysis of governance of hate speech and moderation of platforms, Ahmed and Gillespie (2022) highlight the differential impacts of content moderation policies on some communities and the way certain topics are more likely to be identified as sensitive, while

others are more likely to be misclassified through algorithms. The analysis they conducted shows that moderation is not a one-size-fits-all experience for a platform's users, and that costly moderation is disproportionately borne by marginalized communities who are more likely to talk about violence, discrimination or trauma but are algorithmically similar in these discussions to the harmful content the systems were created to moderate.

This is an insight that directly applies to the study of algospeak. As content moderation is distributed unevenly so are the linguistic changes it engenders, and as a practice it is not necessarily equal along lines of communities of users, algospeak must be understood as such. Algospeak can be the only way a mental health advocate can maintain a lifesaving form of communication, or the only way an activist can record the state's brutality. As in both cases, it is a tension of platform governance that Ahmed and Gillespie (2022) identify: the tension between the platform's wish to regulate harm, and at the same time serve as a space where power over speech is wielded and defied.

Summary and Research Gap

The literature reviewed in this chapter provides a firm and fruitful basis for the current research. Past scholarship has noted the formal characteristics of algospeak, schematized the governance architectures that give rise to it, theorized the critical analytical instruments needed to study it, and shed light upon the differential power dynamics of which it is a part. A sustained critical discourse analysis of algospeak, however, as it functions on X, a platform whose history, culture of moderation, and institutional upheaval in recent years makes it an ideal site to ask questions about it, remains under-explored. This study directly tackles this gap, using the methods of CDA to analyze an X-derived corpus of algospeak practices to explore both the linguistic strategies that users have at their disposal and the power dynamics that they reveal.

Critical Discourse Analysis provides the methodological backbone of this study. As Wodak and Meyer (2022) outline, CDA is not a single unified method but a family of approaches united by a shared commitment to examining the relationship between language, power, and ideology. What distinguishes CDA from other forms of discourse analysis is its explicitly critical orientation: it does not treat language as a transparent medium of communication but as a social practice, through which power relations are produced, reproduced, and, crucially, contested. CDA analysts attend to the ways in which particular discursive choices naturalize certain social arrangements, marginalize certain speakers, and foreclose certain possibilities for meaning-making. Applied to digital communication, CDA encounters both familiar terrain and novel challenges. The volume, velocity, and multimodal complexity of social media discourse require adaptations to conventional analytical procedures, while the algorithmic mediation of digital platforms introduces a new layer of institutional agency that has no direct precedent in earlier discourse contexts (Wodak & Meyer, 2022). Platforms do not merely host discourse; they actively shape it, curating what is visible, amplifying what is profitable, and suppressing what is deemed risky. CDA's attention to the institutional dimensions of discourse makes it particularly well suited to this environment, as it provides the conceptual tools necessary to analyze not only what users say but the structural conditions under which they say it.

Research Methodology

Introduction

The methodological design of this study is a research approach, which is presented in this chapter, along with the methods used to investigate the phenomena of algospeak as a subversive linguistic practice on X. It presents the rationale for choosing a qualitative research design based on Critical Discourse Analysis (CDA) provides information on data collection and corpus construction and describes the analytical framework used for analyzing the data. The chapter also addresses questions of ethical considerations, and the limitations of the chosen methodology. The overall purpose is to show that the methodological decisions are coherent, well-argued, and suitably related to the research questions of the study and its theoretical approach.

Research Design

The research method used in this research is Qualitative. Qualitative methods are suitable when research questions are designed to gain understanding of the meanings, processes and social contexts involved with a phenomenon in order to provide a description of its frequency or distribution (Gorwa et al., 2020). Considering the current study deals with the way algospeak is used as a discursive practice (as a tool for resisting, maneuvering through power, and uncovering the user/platform dynamic), a qualitative design is not only suitable but essential. A quantitative approach is not sufficient to answer both of the research questions and to be rich and interpretive enough.

This study is carried out within the qualitative approach of discourse analysis research design, which is developed based on the theory and method of critical discourse analysis (CDA) as explained by Wodak and Meyer (2022). CDA is a theoretical framework and analytical methodology: it offers conceptual frameworks to understand the link between language and power and sets out procedures for the systematic analysis of texts. Its dual nature makes it ideally qualified for a study that is both formal and social and institutional.

Data Collection

Selection of Platform

Data for this study were collected from X (formerly Twitter). There are several reasons for choosing the site for data collection. The first is that X is a text-based platform, making it especially suited to discourse analysis techniques that focus on the analysis of written texts. Second, and not least, X has been the location of well documented and widely reported algospeak practices, so it is a rich, relevant source of the phenomenon under investigation (Steen & Prahl, 2023). Third, the recent and current moderation instability of the platform, since its acquisition and restructuring in 2022, has been a crucial time to study the platform's moderation and the increasing conditions that foster algospeak.

Corpus Construction

The selection of the data used in this study was based on the purposive sampling method, which is a non-probability sampling method in which the data sample is selected intentionally, not

randomly, according to the criteria for choosing the data sample based on the relevance and significance of the data to the problem of study (Steen & Prahl, 2023). Purposive sampling is best for qualitative discourse analysis, where depth of analysis and theory is the desired objective, not the generalizability.

Known algospeak terms and their typical translations were used to conduct keyword searches of posts in four thematic categories: mental health discourse (e.g., "unalive," "sewerslide"), political speech (e.g., "legal," "bn"), and posts about content moderation (e.g., "shadowban," "algospeak," "getting suppressed") along with emoji-based algospeak. These domains were chosen as they were considered to form distinct communities of users with varying interests in algospeak, which gave a comparative perspective to the analysis. A total of 25 posts were chosen within the corpus which represented a good amount of text data to be analyzed qualitatively but still contained in the study's scope.

Timeframe

The data collected were limited to the posts posted between January 2025 and May 2026. The selected time period was designed to capture algospeak practices after the most salient post-acquisition policy changes by X, so as to capture the current discursive environment on the platform and not a previous iteration of the platform's policies.

Analytical Framework

Critical Discourse Analysis

The basic analytical framework used in this study is the Critical Discourse Analysis (CDA) theory proposed by Wodak and Meyer (2022). CDA is based on the idea that texts are not autonomous, but are created in a particular social, institutional and historical setting, and that their meaning is created by the power relations that exist within the setting. This analysis considers therefore three levels:

- The textual level (the formal linguistic properties of individual posts)
- The discursive practice level (how posts are created, shared, and received in the context of the platform) and
- The social practice level (the wider power relations and ideology that the discourse illustrates and/or reinforces).

Analytical Procedures

On the textual level, each post of the corpus was analyzed for the occurrence of the algospeak strategies used in it. These strategies were coded on the basis of the taxonomy of formal mechanisms formulated by Steen and Prahl (2023): lexical substitution (replacing a flagged word with another), phonetic approximation (changing the spelling of a word to preserve its sound but avoid a keyword), graphic distortion (including the introduction of symbols or punctuation within words), and semantic displacement (using an unrelated word or emojis to communicate a suppressed meaning). The taxonomic coding allowed the systematic identification of patterns within the corpus.

At the discursive practice level, analysis focused on the contextual and inter-textual dimensions of each post, the conversational threads that it was part of, the norms enacted in those threads, and the extent to which algospeak was used explicitly, reflexively, or invisibly in the threads. At the social practice level, the analysis was informed by and adopted from the platform governance scholarship of Gorwa et al. (2020) and Ahmed and Gillespie (2022), which drew on the concept of 'ideological' interpretation of the patterns identified and introduced, linking formal linguistic choices to the structural conditions that give rise to algorithmic moderation practices.

Ethical Considerations

All the data used in this study are publicly available posts on X. There were no private or protected accounts accessed; no personally identifiable information was gathered or stored other than that made public in the public posts. Following ethical recommendations for social media research, the users' username has been anonymized when presenting the data to ensure the protection of individual users' privacy (Zeng & Schäfer, 2021). The study did not include any direct contact with human participants and hence did not require approval of institutional ethics board according to relevant institutional guidelines. However, care was taken in every activity of the analysis, especially those surrounding mental health, self-harm, and other sensitive subjects.

Limitations of the Methodology

There are some methodological constraints that need to be recognized. First, the purposive sampling strategy used for this study is appropriate for qualitative research, and thus the corpus is not representative of algospeak on X as a whole, so that claims are analytic and interpretive rather than generalizable. Secondly, the constraint on the corpus to English-limited posts reduces the cultural and linguistic range of the results, and algospeak could be used in other linguistic situations with other logics. Thirdly, because it is a form of textual analysis, CDA does not access users' intentions or motivations directly, but rather they are inferences based on contextual evidence, not on authorial confirmation. The limitations presented here do not detract from the validity of the analytical statements made in the study within the context established by its research questions, and are consistent with the qualitative design chosen.

Data Analysis and Discussion

Introduction

The analysis is based on a sample of 25 posts gathered from X from 2025 to 2026. The corpus was categorized into four thematic domains: Mental Health Discourse (U1-9), Political Speech (U10-16), and Meta-Discourse on Content Moderation (U17-20), which included an extra set of algospeak posts based on emojis (U21-25). The proposed three-level analytical approach of Critical Discourse Analysis (CDA), textual, discursive practice and social practice, is employed in the analysis of both research questions: How do language strategies that constitute algospeak act as a resistance to algorithmic content moderation, and how do these strategies reflect and reproduce broader power relations between platform governance structures and user communities (Wodak & Meyer, 2022)? This process is explored in each section, where the formal characteristics of the identified algospeak strategies are investigated, as well as their discursive contexts and their ideological value in the context of the platforms' power.

Domain 1: Algospeak in Mental Health Discourse

Lexical Substitution as Protective Encoding

Lexical substitution (replacing the flagged term with one that is phonetically or semantically similar) is most common in the mental health area. The most obvious example in the corpus is *unalive* as in U1:

I'm already sick of this new euphemism for murder, to 'unalive' someone."

In this post, the post itself does not employ algospeak to hide what is not really meant, but rather critically reflects on the birth and spread of the term. The user uses “unalive” as an euphemism, this is a statement that unwittingly further validates the discursive role of the algorithmic language “unalive” by blurring the semantic weight of “kill” or “murder”, just enough to get through algorithmic filters and remain comprehensible for human eyes. This reflexivity with regards to algospeak as a social phenomenon is important in and of itself: the social phenomenon is so normalized that it has caught the attention of average users for meta-commentary (Steen & Prah, 2023).

In U2: a more direct and emotional deployment occurs:

"Life is too good to sit here and depressed tweet so instead I'm going to be submitting my sewerslide note by fax machine."

The appellation *sewerslide*, a phonetic perversion of "suicide", is conveyed in an undertone of black irony. The post's juxtaposition of suicidal thoughts and intentional humor ("by fax machine") is double edged: it contains the message in the text to escape moderation and it protects with humor in the emotional sense. The observation that algospeak often does both, avoiding while also expressing emotion, as noted by Steen and Prah (2023), fits this chain of events. The communicative power of the discourse is deeply unequal, as the user needs to speak in a distorted language to be understood on a public platform, from a CDA perspective (Gorwa et al., 2020). *"From substituting sensitive keywords with phonetically similar words or creative misspellings to using coded language that only in-group members can decipher, algospeak highlights the ingenuity of Internet users in their quest to express themselves freely"* (Bury, 2025).

Graphic Distortion and Numeric Substitution

In Post U4, a more technical approach of numeric substitution is used:

"I am in a big d3pression since my mother passed away and only thing that help me to cope with it is my fandoms. but my sisters r mocking about my obsession... i dont know if they want me d3ad, i can do it asap."

This is because the numbers 3 has replaced the letter *e* in *d3pression* and *d3ad*, which is a calculated graphical distortion (for the purpose of bypassing keyword-based filters). The emotional quality of the post is what makes it analytically interesting: a disclosure of suicidal ideation, not encoded, but rather expressed in the raw, the expression is not delivered ironically, but rather with a sense of urgency; not with somebody else's knowledge, but with the user's, who has presumably

had some prior experience, or heard about it from others, that such expressions are subject to suppression or removal. The communicative urge of the user can't be separated with his/her technical adaptation. This is an example of Ahmed and Gillespie's (2022) concept of the '*uneven burden of content moderation*', where vulnerable users are talking about real psychological distress, and their most pressing communications need to be encoded to be received.

Acronymic and Subcultural Encoding

Posts U5 and U6 demonstrate a less dense form of algospeak: a slang or subcultural vocabulary that has an encoded meaning for those in the community, but is seemingly benign to algorithmic systems. In U5, the term *menty b* (which translates to "mental breakdown" in diminutive form) is paired with the adjective *demure*, a social media term for "modest", to craft a post that feels like an everyday lifestyle post but also recognizes a real mental health experience. Likewise, U6's reference to "edc," which is a short-hand code for talking about eating disorders within certain Internet communities, and a reference to the Electronic Daisy Carnival festival, is a community shibboleth, easily understood by those who are part of the subculture, but untouchable by algorithmic scanners.

From a CDA perspective these are especially important types of subcultural encoding, because they show how "community-internal lexicons" develop, these being special sorts of vocabulary that arise "where there is a shared experience and a shared exposure to moderating pressure" (Steen and Prahl 2023: 137). The algorithmic governance of the platform is not just a suppression of individual posts – it's a creation of alternative semantic systems within user communities.

Symbol Substitution in Sensitive Content

In regard to conversations about sexual violence and abuse, Posts U7 and U8 describe using symbol substitution (replacing letters with typographic characters):

"We should never tolerate r@pe, behe@dings, violence..." (U7)

*"...covered up child S*X AB*USE..."* (U8)

In both instances, the encoding takes place in terms of violence and abuse as described in clearly condemnatory statements. The users are not glorifying or encouraging the acts that they describe, but rather condemning them. But the expected logic of algorithmic moderation — one that could send out those sorts of terms even in spite of the attitude — drives the users to encode even their denunciations. An example that is very revealing of the ideological constraints of keyword-based moderation: it cannot read sentiment, it can read only surface form, and thus the communicative load placed on anyone who condemns violence will be the same for those who might celebrate it (Gorwa et al., 2020). The outcome is a discursive space in which moral clarity has to be conveyed in a way that is linguistically obscure.

Domain 2: Algospeak in Political Speech

Graphic Distortion in Contested Political Discourse

In the political domain, the graphically distorted terminology is consistently applied to politically sensitive terms. Post U10 is an ironic use of the standard orthography against algospeak:

"Say it with us... ILL. LE. GAL. ILLEGAL. And no. Stop it."

The post explicitly opposes the algospeak encoding of law, by spelling it out in capitals letter by letter — a performative affirmation of the right to use language not encoded. The surprise use of this meta-discursive movement is that it affirms the currency of the encoded form: the user knows that legal is a variant of algospeak and resists it. Algospeak and its rejection are both discursive acts, as illustrated in this post, as part of discursive struggles with political speech and the legitimacy of immigration discourse, as described by van Dijk (2008) (Wodak & Meyer, 2022).

Post U11: This shows typical distortion in a political context:

*"G*ns DO NOT Kill People, PEOPLE KILL PEOPLE... US peeps owned G*ns since 1789."*

The asterisk in *g*ns* is a substitute that indicates the user's understanding that gun-related material is monitored algorithmically on X. The encoding enables the customer to immerse themselves in a politically charged discussion — gun rights — while maneuvering the platform's systems of moderation. Notably, this user's argument is a politically conventional one in favor of gun rights, which implies that there is no partisanship attached to the algospeak, and users have been known to use it on both sides of the political spectrum, where they see the threat of algorithms (Gorwa et al., 2020).

Religious and Geopolitical Encoding

Posts U15 and U16 demonstrate algospeak deployed in geopolitically sensitive contexts, specifically around discussions of Palestine and Islam:

"...voice support for P@lestin!ans, ending with the slogan 'Free P@lestin!ne.'" (U15)

*"I'd heard of J*ihad before, but somehow I wasn't previously aware of their sacred duty..." (U16)*

For U15, the encoding of 'Palestinians' and 'Palestine' is an indicator that a user knows that pro-Palestinian content has been documented to be suppressed on major social media platforms. The use of @ as a and ! as an i is a subtle but intentional manipulation of the images' graphical form in order to maintain readability by the human eye and avoid being filtered by keywords. The term religious and political Jihad, which has complex connotations, is an illustration of algorithmic moderation systems flagging terms that have religious extremist associations, even when they are used in an analytical or educational context, in U16. In both posts, the sentiment of the content is targeted with moderation measures that are disproportionately applied to communities that have been algorithmically tagged as "sensitive" to their topic, irrespective of the purpose and tenor of the speech. In both posts, the content is flagged with moderation actions that are more likely to

target communities that are algorithmically labeled as “sensitive” to their topic, regardless of the purpose or tenor of the speech.

Direct Reference to Censorship

Post U14 is analytically distinctive in that its subject matter is itself the experience of censorship:

"Conservative are still being censored on this platform Elon You need to stop this abuse of being censored... Retweet this if anyone gets to see it so Elon can see what's going on."

Notably, this post is not an algospeak-encoding political grievance (that is, it does not use a coded way of saying the word). But, because of its appeal ("if anyone gets to see it") it also acknowledges suppression, thereby communicating its own communicative vulnerability it is protesting. The post is an expression of a transparency act: a clear statement that one's speech is being censored and that it is stated in a way that is exactly how one thinks it would not be heard. This self-presentation of platform power represents one of the most straightforward examples in the corpus of platform power and platform governance structures (Gorwa et al., 2020).

Domain 3: Meta-Discourse on Content Moderation

Reflexive Awareness of Algorithmic Governance

The domain of the meta-discourse is the most analytically rich part of the corpus due to the fact that posts in this domain are explicitly concerned with the experience of algorithmic government. After U17, directly refers to the platform mechanism that the user has encountered:

"So apparently sharing a grok imagine video doesnt add the made with AI tag. So I deleted it. Tried resharing and now it shadowbanned me from resharing it. UGH"

Shadowbanned, a word that has become part of the tech jargon of the subculture and now is part of the jargon of the platforms themselves, is an indicator of advanced platform sophistication in knowing how the platform is moderating information. The User's Frustration ("UGH") is channeled not at a person making a decision but at an algorithmic system that operates in a mysterious way and whose results feel random. This is an example of Gorwa et al.'s (2020) understanding of the opacity of algorithmic moderation - that the opacity of algorithmic moderation is itself a form of power, users cannot appeal, negotiate, or even understand the decisions that govern their communicative experience.

Perhaps the most explicit, in the entire corpus, is the observation made in Post U18:

"I can't believe people on tiktok really chose 'sewerslide' as the algospeak replacement for 'suicide', sounds humorously offensive while also reminding me of the scene from Mouse Hunt..."

Algospeak is the name given to this phenomenon, it is evaluated aesthetically and ethically (humorously offensive), and it is placed within a cross-platform ecology (TikTok/X). It is both a full reflexive metadiscourse: a user of an online platform, reflecting critically on the language practices which the platform has generated. Significant from a CDA perspective is that this

reflexivity indicates that algospeak has gained its own cultural visibility, and now can become an object of public discourse, rather than a secret survival strategy as Steen and Prahl (2023) noted.

Posts U19 and U20 further the domain of meta-discourse, offering commentaries on the political aspects of algorithmic curation:

"Has anyone noticed how every post supporting Iran is getting suppressed yet?" (U19)

"Okay, I am overseas so the algorithm hates me." (U20)

U19's conspiracy of suppression, geopolitically inflected content moderation, told as a coordinated experience (Has anyone noticed? The post welcomes expressions of solidarity from users who hold the same opinion; that one's right to speak is being politically silenced. U20, in contrast, sees the algorithmic hostility as a geographic rather than content-based issue, and makes the power of the algorithm "readable" as a relationship with the algorithm ("hates me"). Both posts present the users' efforts to make sense of opaque moderation systems, which they attribute an intention to, whether that be political, geographic, or otherwise, but, formally, are automated processes.

Domain 4: Emojis-Based Algospeak

There is a type of algospeak where emoji are used as the main encoding tool, which includes posts U21 through U23. The ✂ emoji is used in U21 in conjunction with words about military strikes; the emoji in U22 with explicit mention of street violence; and the emoji in U23 when the post is about conflict, as a component of a political message. The emoji in each of these instances is not meant to be a decorative addition to the words, but rather a semantic carrier: it communicates a meaning of violence, conflict or threat in a register of visual symbolism that is less well suited to the task of filters based on keywords.

The examples in these posts depict the semiotic ingenuity of algospeak, where there's no lexical substitution available or at all, users resort to alternative semiotic modes. The shift of communicative content from verbal to visual (word to emoji) is an adaptive response to the specific technical architectures of automated moderation systems that have historically been based on keyword detection in text rather than on multimodal semantic analysis (Gorwa et al., 2020). The outcome is a means of code-switching communication that works exactly between the reach of the human and the machine.

Summary of Findings

The following summary of the findings of the corpus analysis is related directly with the research questions of this study. Algospeak on X works on a variety of formal strategies that are all customized to the context of the moderation and the users it targets, such as lexical substitution, graphic distortion, numeric substitution, acronymic encoding, subcultural lexicons, and emoji displacement. Second, such strategies are also communicative resistance strategies, or strategies by which users can sustain and carry meaning that platform governance systems often attempt to marginalize, divert, or obscure. Third, the presence of algospeak on different domains shows that the burden for encoding is unevenly distributed and that those who are most aggressively

moderated by the platform are also the ones who speak about mental health, political minorities, and geopolitically-sensitive topics. Finally, the meta-discourse domain displays high user awareness of the algorithmic systems that shape their speech, indicating that resistance with algospeak is not primarily an automatic adaptation, but rather, in many cases, a politically conscious act.

Conclusion and Recommendations

Conclusion

The purpose of this study was to explore algospeak on X from a Critical Discourse Analysis perspective, addressing two research questions: 1) What strategies are used in the production of algospeak that enable it to act as resistance, and 2) What power relations in the process of language production does algospeak reveal and reproduce? The analysis of 25 real posts gathered from X from January 2025 to May 2026 gave insight that directly answers both questions.

Six main linguistic strategies were found to make up algospeak on X, according to the first research question: lexical substitution, phonetic distortion, graphic distortion through symbol and numeric replacement, acronymic encoding, subcultural encoding, and emoji based semiotic displacement. These are different ways of dealing with the technical infrastructure of algorithmic content moderation so as to maintain intent for humans to read and act on the content, yet avoid detection by machines. Significantly, these strategies are not simply a form of avoiding moderation, but are ways of communicating resistance that establish a claim to the right to speak on topics that are otherwise being denied or limited by the platform's governing structures (Steen & Pahl, 2023).

To address the second research question, the analysis identified a range of algospeak practices that are deeply rooted in a structural imbalance of power between the governance organizations of the platform and the communities that use it. Algorithmic moderation systems operate in the dark, are context and sentiment agnostic, and are not applied even-handedly, with users of mental health, political minorities and geopolitically sensitive topics being most likely to be the targets of their regulation (Gorwa et al., 2020). The meta-discourse domain of the corpus also showed that the users have a very well developed understanding of these dynamics and that in many cases algospeak is not simply an automatic response but is a conscious political stance of resistance.

Recommendations

For Platform Governance

The context-sensitive moderation architecture of content moderation systems on X and other similar platforms should be accelerated to take a more holistic approach at evaluating the sentiment, intent and communicative context of flagged content, rather than primarily for keyword detection. The corpus offers clear evidence that keyword-based systems on the one hand burden users who object to violence, and on the other hand burden users who embrace violence, and that the same is true of users who talk about mental health in crisis versus those who use vulnerable discourse. Platforms should commit to investing in moderation tools that can differentiate between types of sensitive language use such as condemnatory, educational, and supportive, from harmful content (Gorwa et al., 2020). More transparency around moderation decisions, such as clear

explanations about why content has been suppressed, would also be a meaningful step toward reducing the uncertainty that users experience when they use communication platforms, which leads them to use 'algospeak'.

For Digital Literacy Education

Based on the results of this study, it is recommended that algospeak literacy, which is the understanding of the process and purpose of how users communicate through text on social media should be introduced into the digital literacy curriculum at secondary school and tertiary level. It is important for teachers to provide students with the needed critical skills to see platform governance not as an impartial technical process, but as an exercise of institutional power that has real consequences for communicative freedom, especially for marginalized communities (Ahmed & Gillespie, 2022).

For Future Research

This study could be expanded in three ways in future research. A cross-platform comparative study of algospeak on X, TikTok, and Instagram would reveal how distinct structures of moderation result in different linguistic strategies of adaptation. Second, multi-lingual corpus research would help determine if there are different formal logics underlying algospeak in other languages than English. Third, longitudinal studies on the development of the individual algospeak terms over time could help to clarify the dynamic relationship between updates to the moderation system and adaptation of users' language (Zeng & Schäfer, 2021).

Algospeak isn't just a curiosity in language. It is a documentary record of the conflict between the institutional power and the communicative power in the digital age. Each misspelled vowel, each substituted symbol, each reappropriated emoji in the data set analyzed by this study is a protest from a user who does not want to be silenced, a protest that is small, persistent, and exceedingly creative, a protest that is a right to speak. These coded conversations are not just about language; they are about power, resistance, and the persistence of every human being to be heard

References

1. Ahmed, W., & Gillespie, M. (2022). Hate speech, content moderation, and the governance of social media platforms. *Journal of Information Technology & Politics*, 19(3), 245–261. <https://doi.org/10.1080/19331681.2021.1994130>
2. Bury, B. (2025). Decoding algospeak: Unveiling the patterns of linguistic evolution in the digital age. *Academic Journal of Modern Philology*, 27, 37–51. <https://doi.org/10.34616/ajmp.2025.27.3>
3. Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society*, 7(1), 1–15. <https://doi.org/10.1177/2053951719897945>
4. Mahl, D., Schäfer, M. S., & Zeng, J. (2023). Conspiracy theories in online environments: An interdisciplinary literature review and agenda for future research. *New Media & Society*, 25(7). <https://doi.org/10.1177/14614448221075759>

5. Wilson, Richard A. and Land, Molly, "Hate Speech on Social Media: Content Moderation in Context" (2021). Connecticut Law Review. 449. https://digitalcommons.lib.uconn.edu/law_review/449
6. Wodak, R., & Meyer, M. (Eds.). (2022). *Methods of critical discourse studies* (4th ed.). SAGE Publications.

Appendix

Corpus

Date Posted	Actual Text	Algospeak	Domain	URL
29-Apr-25	I'm already sick of this new euphemism for murder, to "unalive" someone.	Unalive	Mental Health	https://x.com/CatholicSquid/status/1917211266436731067
17-Apr-25	Life is too good to sit here and depressed tweet so instead I'm going to be submitting my sewerslide note by fax machine	sewerslide	Mental Health	https://x.com/0M3NSDEAD/status/1912898342968770989
26-Apr-25	Norf robbed there. Joke the umpires didn't see that blatant block	Norf	Mental Health	https://x.com/timgg1971/status/1916055130463735881
25-Feb-25	i am in a big d3pression since my mother passed away and only thing that help me to cope with and handle it is my fandoms. but my sisters r mocking about my obsession over the things that actually make me happy. i dont know if they want me d3ad, i can do it asap.	d3pression, d3ad	Mental Health	https://x.com/bartyhasissues/status/1894402865185407396
11-Apr-25	had a quiet demure menty b in my work bathroom and now I'm feeling a little better ☐	menty b, demure	Mental Health	https://x.com/bucksلاسagna/status/1910449880704967111
29-Apr-25	Vegas this weekend and then back for edc ☹️ WHERE DOES IT END	edc	Mental Health	https://x.com/LarinEricka/status/1917279606673137981
31-Aug-25	We should never tolerate r@pe, behe@dings, violence, church burnings, or child marriages. That only enables the atrocities to continue.	r@pe, behe@dings	Mental Health	https://x.com/Eyes4Eternity/status/1961928984280625358
20-Jul-25	In 2018, Francis wrote an open letter	S*X, AB*USE	Mental Health	https://x.com/pinkcakecream/status/1946662488847663380

	defending cardinal wuerl who set up hush money and covered up child S*X AB*USE and had to resign. EF the catholic church and how they defend toxic priests over children.			
25-Jul-25	she's so weird... first tory now that c*ke addict.	c*ke (cocaine)	Mental Health	https://x.com/icarestan/status/194848855548000297
7-Jun-25	Say it with us... ILL. LE. GAL. ILLEGAL. And no. Stop it. Stop spending our money on everything BUT Texas!	le*gal	Political Speech	https://x.com/WWiilie/status/1931212469608780153
29-Aug-25	MENTAL ILLNESS = School Shootings G*ns DO NOT Kill People, PEOPLE KILL PEOPLE School Shootings started apx 30 yrs ago & US peeps owned G*ns since 1789	g*uns	Political Speech	https://x.com/SalzOld/status/1961208366732017952
28-Dec-25	How bout that MR "I defend Christian" TRUMPO?!? If you REALLY CARE that much abt Christians.. IF CHRISTIANS B*N KILLED R THE REAL REASON YOU ATTACKED NIGERIA.. then WHY are you NOT retaliating AGAINST ISRAEL that KILLED SO MANY CHRISTIANS??? Hmmm.. so what's the truth ole Trumpo	b*n	Political Speech	https://x.com/Miriamschiro2/status/2005159280253784446
29-Dec-25	"Some guys are thr€atening" "Some guys are ready for w@r" "Blow my Trumpet oh God"	thr€atening,w @r	Political Speech	https://x.com/SPYNG247/status/2005692610955120936
31-Dec-25	Conservative are still being censored on this platform Elon You need to stop this abuse of being censored. I'm not the only one speak out Conservatives let him know we're still being censored. Retweet this if anyone gets to see it so Elon	censored	Political Speech	https://x.com/MaryKno38656667/status/2006118466974076990

	can see what's going on.			
15-Sep-25	Jewish Actress Hannah Einbinder, winner of the Emmy for Supporting Actress in a Comedy Series for Hacks, used her first acceptance speech to voice support for P@lestin!ans, ending with the slogan "Free P@lest!ne." #HannahEinbinder #Emmys2025 #Hollywood #Entertainment #Emmys	P@lest!ne	Political Speech	https://x.com/Islamic_2_0_2_2/status/1967652886349029602
30-Nov-25	I was recently watching a docudrama about the fall of Constantinople. It was the first time I'd heard that the Q.uran had clear instructions for how to invade a city, how to give an ultimatum and discuss terms of surrender with the invaded, and so forth. I'd heard of J*iihad before, but somehow I wasn't previously aware of their sacred duty to encroach on territory. Rise of Empires: Ottoman (2020) on Netflix	J*iihad	Political Speech	https://x.com/Dolly_gale/status/1994943247995781524
31-Mar-26	So apparently sharing a grok imagine video doesnt add the made with AI tag. So I deleted it. Tried resharing and now it shadowbanned me from resharing it. UGH	shadowbanned	Meta-Discourse on Content Moderation	https://x.com/graykevinb/status/2038721273740628378
24-Feb-26	i can't believe people on tiktok really chose "sewerslide" as the algospeak replacement for "suicide", sounds humorously offensive while also reminding me of the scene from Mouse Hunt where the deceased father's body flies into the manhole cover	algospeak	Meta-Discourse on Content Moderation	https://x.com/hyperdiscogirl/status/2026110021902197028
27-Mar-26	Has anyone noticed how every post	getting suppressed	Meta-Discourse on	https://x.com/success1906/status/2037369978114994619

26-Mar-26	supporting Iran is getting suppressed yet? Okay, I am overseas so the algorithm hates me. I am going to bed. Buy my books, thanks. Sincerely, A tired traveler	Algorithm hates	Content Moderation Meta- Discourse on Content Moderation	https://x.com/AuthorTRobare/status/2037239354385392044
------------------	---	--------------------	---	---
